Turning the tide against TB: Remaking ineffective host defenses into mechanisms for tuberculosis control

Turning the tide against TB: Remaking ineffective host defenses into mechanisms for tuberculosis control

A dissertation presented

by

Yanjia Jason Zhang

to

The Division of Biological Sciences in Public Health

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Immunology and Infectious Diseases

Harvard University
Cambridge, Massachusetts

March 2013

Turning the tide against TB: Remaking ineffective host defenses into mechanisms for tuberculosis control

Abstract

Most antibiotics, including the drugs currently used for treating tuberculosis (TB), were first discovered as molecules that inhibit bacterial growth in laboratory culture conditions and later translated to infection models and clinical use. *Mycobacterium tuberculosis* (Mtb) has evolved specifically to survive in its human host, and it is in this infectious context that new drugs need to work. The host environment is characterized by a multitude of antimicrobial defenses induced by the immune system, and we can leverage these defenses to kill Mtb *in vivo*. Mtb employs a diverse set of responses to survive host defenses. By blocking these responses, we can make Mtb more susceptible to host immunity, turning these previously impotent defenses into effective strategies of immune control.

In this dissertation, I characterize tryptophan biosynthesis as a bacterial pathway required to combat the specific CD4 T cell-mediated defense mechanism of tryptophan starvation. Armed with this understanding, I searched for and found small molecule inhibitors of Mtb tryptophan biosynthesis with potency in the micromolar range in media lacking tryptophan. The molecules worked even better in a macrophage model of Mtb infection, and had remarkable synergy with interferon-γ. In macrophages, interferon-γ and the tryptophan biosynthesis inhibitor blocked worked in conjunction to kill Mtb, the former catabolizing

exogenous sources of tryptophan and the latter blocking endogenous tryptophan synthesis.

The findings in this dissertation show that understanding how Mtb genes work in the infectious context can help identify good *in vivo* drug targets. Furthermore, we begin to scratch the surface of failing host defenses that can be revived to effectiveness with the inhibition of the right host-adaptation pathways in Mtb. Essential genes will continue to serve as a good repository of potential drug targets—and to that end, Chapter 2 describes a renewed assessment of Mtb essential genes—but by expanding our understanding the genetic requirements of bacterial survival against immune system components (Chapter 3), we can identify new therapies that help our immune system kill Mtb (Chapter 4).

TABLE OF CONTENTS

Acknowledgements

There are points in our scientific careers when our mentors advise us to use the singular first person, “I.” I was told to use “I” during my preliminary qualifying examination. Post-docs are told to use “I” when giving job interview talks. But it is awkward. “We” feels much more appropriate because it is more appropriate; scientific efforts are rarely lonely. It was awkward, then, when I typed the title page of this thesis and only my name appeared. Like the select times when we are told to use “I,” the single-authored dissertation reflects custom, not reality. This dissertation was a collaborative effort, and thankfully so. I could not have done any of it without the help, support, mentorship and friendship of many I will now thank. More importantly, I would not have wanted to do it without them.

I started the MD PhD program in the summer of 2007 with 4 New Pathway classmates whose solidarity over the past six years has been invaluable. I drove up with of them, Luciano Custo-Grieg, when we first moved to Boston from New Haven. We attempted to drive that U-Haul truck onto Storrow Drive. The truck was too tall for Storrow, and we eventually had to back up and take the long way around. It was a fitting start for us. Gilad Evrony, Joe Bell, David Konieczkowski, and Luciano, we’re still taking that long way around to doctor-hood, but like driving backwards on a Storrow onramp, it’s initially harrowing but ultimately the most thrilling path. Thanks for journeying with me. And a special thanks to David, my roommate of 2 years, groomsman, and source of late night sleep-talking comedy.

That same summer I rotated with the Rubin Lab and immediately loved it. It took me a while to convince myself that I wanted to work with bacteria, but it took no time to convince myself to want to work with the Rubin Lab. Thanks to that first crew of graduate students I overlapped with, Jeff Murry, Sloan Siegrist, Mary Farrow and Mike Chao, whose obvious joy helped get me excited about starting at the bench. When I returned two years later to start full time, only Mike amongst the graduate students remained, but the lab's atmosphere was the same. Fun, excited, close-knit, full of teaching; all attributes that reflect Eric Rubin, our fearless leader. I'll get back to him in a bit.

I want to thank Babak Javid, whose active mind generated about 75% of lab chatter—from the scientific to the frivolous, for his mentorship, teaching, and constant concern for me and the other students in the lab. Jun-rong Wei was another post-doc stalwart. She is a giant reservoir of knowledge in a small always-smiling package, and she never says no when given an opportunity to teach. Jun, watching you help all of us by day and work on your own projects by night was an inspiration. We are all blown away by your selflessness; the lab would not have run without you in those years. Many, many thanks go to all the post-docs: Magnus Steigedal, for your humor and calm and Johannes soccer stories; Flavia Sorrentino, for your kindness and warmth and your trips to Europe which always yielded the best snacks; Alissa Myrick, for buying US a cake and encouraging US to party on YOUR birthday; Giovanna Poce, for your chemical

expertise and your quick laugh; Cara Boutte, for making it possible to find things in our lab; Hesper Rego, for teaching us all what microscopes do, and for your artistic science and scientific art; Amy Barczak, for your sane presence in a crazy lab, and giving our lab another person vaguely interested in the host. Thanks also to Brian Schuster, our outdoorsman extraordinaire and Swiss army knife of a technician. Thanks to Kristi Guinn, Jess Pinkham and Emmy Dove, who together managed our lab and us, a terribly unmanageable crew.

To my fellow graduate students: our elder statesman, Mike Chao, who despite his very best efforts to be discouraging and negative created a tremendously encouraging and positive experience for us younger grad students; Mandy Martinot, without whom none of us would know what to do with animals, both experimental- and pet-purposed; Karen Kieser, always happy and armed with a sonorous belly laugh that exemplifies the fun we have in the lab; Annie Park, who will take the mantle forward when all the above leave; and finally, to Ravi Raju, who joined the lab with me and was a constant companion through thick and thin. Your friendship and wisdom was a buoy throughout our years in the lab. It was a winding path that went from the difficult decision not to join you in Baylor to the decision to partner up in the same lab. I'm glad we ended up in the same place, and hope we end up in the same place again in the future.

Special thanks also to all those with whom I had the pleasure of working together on projects. To Andrej Trauner, it was fun having someone else who preferred

the computer to the bench. Your relentless positivity and your ability to pick up on those difficult quantitative questions made you an awesome partner. But for the record, after 3 years in America, you still can't say "awesome" like an American. To Robin Basu Roy, I'm glad that my poor choices on your behalf didn't drive you completely away from science, and I hope to keep running into you on the Great Wall of China or in Africa or wherever you end up fighting diseases for sick kids. To Justin Pritchard and Jackson Marakalala, you guys have been amazing to have in the lab, and your fresh perspectives provided a huge jolt of energy when I thought I'd be petering out.

If I only had the opportunity to work with members of the Rubin Lab, it would have blessing enough. However, over the past three and a half years in the lab, I got to work with so many others who expanded our understanding of the biology of bacterial infections. Eric Rubin is a naturally collaborative scientist, and I am grateful for his leadership in pushing me into some truly excellent collaborations. In Texas, thanks to Tom Ioerger for your guidance in all things bioinformatics. Also to Joseph Mire, Michael DeJesus Manchi Reddy, Deeann Wallis and Jim Sacchettini, who were all wonderful to work with. To Kirti Sharma and Matthias Mann in Munich, for your generosity and hospitality, and your continued partnership. To Sam Behar and Alissa Rothchild, without whom I would have been completely lost in the swamp of TB immunology. To Curtis Huttenhower, who taught me how to make sense of numbers. To Chris Sassetti, who taught me TraSH and who never stopped thinking about how to make it work better. To

Sarah Fortune and her wonderful lab—it was wonderful to partner with you all so closely. To Abe Brass and Thomas Pertel, who helped me with that fun diversion into host biology. To Veronique Dartois, without whom our small molecule could never have gotten into mice.

And now to the head honcho. I first met Eric Rubin at a fancy event during the MD PhD revisit weekend. It was a wine and hors d'oeuvres event in the most pretentious of our medical school buildings. The admitted students were all dressed up, as were the professors. Eric was in a slightly ratty jacket covering an even rattier T-shirt (the T-shirt description I'm making up… but it is a fair extrapolation of Eric Rubin fashion circa 2007) capped off with a Seattle Mariners hat. He was the most approachable of all professors I met, and his candor and humor immediately disarmed all the things I hard heard about Harvard being impersonal and cold. Through the next two years, Eric helped me identify exciting professors to rotate with, all of whom worked on HIV immunology, my main scientific interest. I had little interest in bacteriology, but over those two years of being counseled by and learning from Eric, I developed a huge interest in working in his lab. I joined without knowing what to work on, but supremely excited about the PI I'd be working with. It was a perfect decision. Eric was as approachable during my time in the lab as the man in the Mariners hat, even as indeterminate factors added collars and buttons to his wardrobe. His mentorship in career decisions continued, and I learned how fun it was to work scientifically with a man whose sharp intellect and creative mind I will always have trouble

keeping pace with. Most important, he is kind, gracious, and caring. He is the complete package as a mentor, and I am extremely grateful for the opportunity to have been in his lab.

I want to thank my other mentors at the medical school and the school of public health. My medical school professors were both teachers and inspirations, teaching us and getting us excited about everything from biochemistry to social medicine. I want to thank especially Joe Rhatigan, Joe Tucker, Buck Strewler, Sam Behar, Julian Seifter, Nir Eyal, Paul Farmer and Jim Kim. I want to thank my wonderful lab rotation mentors as well: Thumbi N'Dungu, Victoria Kasprowicz, Bruce Walker, David Kauffman, Dan Barouch, Fabio Martinon and Laurie Glimcher. The members of the BPH program have been tremendously supportive: Ruth Kenworthy, Holly Southern, and our director Marianne Wessling-Resnick. The MD PhD program has kept track of us throughout the years and provide excellent support especially at the awkward transitions. I thank our former director, Stephen Blacklow and our current director, Marcia Goldberg, as well as the host of administrative mavens who patch a dispersed network of training options into a unified program: Linda Burnley, Yi Shen, Robin Lichtenstein, Amy Cohen, Jennifer Deangelo and Paula Harwick. I want to thank Barry Bloom and Rich Losick, who I had the pleasure of teaching with and learning from in LS120. Their enthusiasm for science and the applications of science to global health threats was exciting to watch week in and week out. The guidance from my amazing dissertation advisory committee was invaluable,

thanks to Marcia Goldberg, Curtis Huttenhower, Chris Sassetti and Sam Behar. I want to thank all my pre-grad school mentors as well, who inspired me to pursue science and medicine: Gay Crooks and Kim Payne at Children's Hospital LA, Stu Orkin and Jon Snow at Boston Children's, Eric Herzog and Diane Krause at Yale.

The start of my thesis work coincided with our joining a new church plant in Coolidge Corner. The community at City on a Hill Church has been a tremendous support and a source of great teaching and fellowship. I want to thank especially the various members of our community group for the jolt of joy and wisdom that floods into our apartment each Wednesday evening.

Finally, and most importantly, I want to thank my family. My parents have been a source of support and inspiration all of my life, and Mom and Dad were instrumental in nurturing my interest in biology. My wife, Veronica Zhang, is a constant source of joy and love in my life. Your support in the easy times and hard is such an encouragement, and it still amazes me that your willingness to keep me company on the late nights in lab has yet to wane. This is only the beginning of our journey together, and it the main excitement of my life to know that I get to journey with you for the decades to come. Finally, I can't say this an better than Jeff Murry did, so I'll just quote him: "God provided the path that I am on and the motivation that I to follow it."

Thank you.

Chapter 1.

Overview and Introduction

This chapter contains one manuscript that is in press as a review:

Zhang YJ, Rubin EJ. Feast or Famine: The Host-Pathogen Battle over Amino Acids. *Cellular Microbiology*. In press.

Section 1. 1: Thesis Overview and Attributions

Overview

For most of the 20th century, Tuberculosis (TB) was a disease on the decline. Improved sanitation and living conditions had helped put TB on a downward trajectory, and the advent of semi-effective vaccine and nearly fully-effective antibiotics made eradication of the disease seem possible, if not imminent. However, two major factors crept in around the 1980s that changed reversed the course of the global Mtb epidemic. First, drug resistance became an issue, and it soon became appreciated that there was resistance against all available drugs, and that certain strains harbored multiple drug resistance alleles. Second, the HIV epidemic created huge populations of people who had an increased risk of getting TB. Today, TB is once again one of the preeminent threats to global health, infecting about 2 billion people, causing disease in about 10 million each year, and killing about 2 million this past year.

Discovering new drugs for TB, then, is one of the main global health priorities. Current drug regimens are able to both treat active disease and prevent future disease by drastically decreasing the risk of reactivation. However, even the shortest treatments require daily dosing of 2-4 drugs for 6 months, and many drug resistant strains necessitate up to 30 months of treatment. Accelerating the treatment process would be a tremendous aid to global anti-TB efforts. Furthermore, while multiple drug resistant (MDR) cases are still treatable by a combination of available drugs, extensively drug resistant (XDR) and totally drug

resistant (TDR) strains are becoming alarmingly prevalent, and treatment of these bacteria desperately is in desperate need of new drugs and new drug targets.

Discovering new drugs could be aided by the identification of mycobacterial processes to target. Traditionally, drugs have been discovered by screening small molecule libraries in laboratory culture conditions to identify small molecule inhibitors of essential growth processes. Identifying these essential processes, then, is a crucial task in determining targets that would be good for drug design. To that end, Chapter 2 of this thesis describes a process by which we determine the requirement for growth of just about every gene in the genome. We introduced a transposon into a pool of bacteria, such that all bacteria in the pool have a single transposon disrupting a single genomic site. We then employed deep sequencing technology to map the insertion sites of each transposon in the library. Making the assumption that genomic regions lacking insertions represent genes required for optimal growth *in vitro*, we searched the genome for these regions with an underrepresented number of reads, and were able to identify genes, individual protein-domain-encoding regions, small RNAs, and other intergenic regions that were required for growth.

Proteins required for in vitro growth are not the only proteins worth targeting for drug treatment. Mtb, after all, lives primarily in the host, and the host is a very different environment than laboratory cultures. To that end, Chapter 3 describes

the studies that helped us determine the genes required for growth in a mouse. We used the same transposon library, comparing the surviving pool of bacteria retrieved from the mouse to the library with which we infected the mice. We were also interested specifically in the mycobacterial determinants of surviving elements of the immune system. CD4 T cells are one of the most important groups of adaptive immune cells. They are important for limiting bacterial growth one month after infection, but even at their fullest induced activity in mice, Mtb is able to survive the effects of CD4 T cells We infected our library into both wildtype mice and mice lacking CD4 T cells to find the mycobacterial determinants for surviving CD4-mediated stress.

Amongst many other things, we found that CD4 T cells induce the requirement for Mtb to make its own tryptophan. This suggested the CD4 T cells have a strategy to decrease the amount of exogenously available tryptophan to Mtb, and the first half of Chapter 4 is dedicated to the elucidation of this CD4-mediated starvation strategy. Using a cell culture model of Mtb infection and CD4 T cell help, we determined that CD4 T cells, through the secretion of interferon-γ, induces an intracellular enzyme, indoleamine-2,3-dioxygenase, that depletes intracellular tryptophan, making it necessary for Mtb to produce its own tryptophan during infection.

This was an interesting pathway, but a futile one in Mtb. We described a starvation pathway that would have killed certain tryptophan-auxotrophic

pathogens, but could not kill Mtb because of its ability to synthesize tryptophan. On the other hand, it also represented a potential point of therapeutic intervention. If we could inhibit tryptophan biosynthesis, we would be able to negate Mtb's adaptation to tryptophan starvation, and in the process turn an ineffective host response into an effective mechanism of TB control. The second half of Chapter 4 is devoted the discovery of such an in inhibitory compound. We show that this small molecule is able to kill TB in media lacking tryptophan. We show also that this small molecule is able to kill Mtb in macrophages. Because we showed that tryptophan biosynthesis was demanded by CD4-mediated mechanisms, we predicted that a small molecule inhibitor targeting tryptophan biosynthesis would make Mtb more susceptible to CD4 T cells. Indeed, when treated with both CD4 T cells in co-culture or with interferon-γ, the small molecule does kill Mtb at a much higher rate, demonstrating its synergy with these mechanisms of immune activation.

In summary, this dissertation identifies genes required for growth in vitro, genes required for growth in vivo, and the subset of the latter that are required for surviving the CD4 T cell response. By identifying tryptophan biosynthesis as part of the indispensible response to CD4 T cells, we note that this biosynthetic pathway is ripe for chemical inhibition to kill Mtb in conjunction with the immune system. Finally, we identify a small molecule inhibitor of tryptophan biosynthesis and show its synergy with the immune system in a macrophage culture. As this

dissertation is being presented, the small molecule is being tested for efficacy in mice, in hopes that it will be able to kill Mtb in an established mouse infection.

Attributions

In addition to the general introductory statements in Sections 1.1, this Chapter is also composed of two manuscripts, both written by Eric Rubin and myself. The first manuscript is currently in press at *Cellular Microbiology* as a microreview. I wrote this review while in Babak Javid's lab in Tsinghua University in Beijing, and am grateful for his input and his motivation during that time. Cara Boutte, Hesper Rego and Kristi Guinn, members of the Rubin Lab also helped me revise this manuscript. The second manuscript is a draft of a review that we intend on submitting as a review article.

Section 1.2: The Host-Pathogen Battle over Amino Acids

Overview

Intracellular bacterial pathogens often rely on their hosts for essential nutrients. Host cells, in turn, attempt to limit nutrient availability, using starvation as a mechanism of innate immunity. Here we discuss both host mechanisms of amino acid starvation and the diverse adaptations of pathogens to their nutrient-deprived environments. These processes provide both key insights into immune subversion and new targets for drug development.

Attributions

The following section is in press as a review article at Cellular Microbiology. Thanks to Cara Boutte, Hesper Rego and Kristi Guinn for help in reviewing and revising this manuscript.

Feast or Famine: The Host-Pathogen Battle Over Amino Acids

Yanjia J. Zhang[1], Eric J. Rubin[1*]

[1]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA

*To whom correspondence should be addressed:

Address: 4 Blackfan Circle, HIM 1048, Boston, MA 02115

Phone: 617.432.3335

Fax: 617.738.7664

Email: erubin@hsph.harvard.edu

Running Title: Host-Pathogen Battle over Amino Acids

Word count: 3,934

Introduction

Successful bacterial pathogens exploit their hosts to support their own survival. Pathogens rely on their hosts for nutrients necessary for survival, injuring the hosts in the process. For their part, the hosts are hardly gracious. In fact, the ability to keep would-be pathogens away from the nutrients turns out to be a key part of host defense. This interplay exists for all essential nutrients, including carbon, nitrogen, and transitional metals(Eisenreich *et al.*, 2010, Rohmer *et al.*, 2011, Hood *et al.*, 2012, Skaar, 2010).

For many potential pathogens, amino acids are critical nutrients. While many organisms can synthesize their own amino acids, others must scavenge them in order to make proteins. Additionally, many bacterial pathogens are able to use amino acids as a carbon source, some even as their principle carbon source(Venugopal *et al.*, 2011, Eylert *et al.*, 2010, Molofsky *et al.*, 2004, Wieland *et al.*, 2005). Accordingly, mammalian hosts have evolved mechanisms to starve bacteria of amino acids. And, to counter this, bacteria have a diversity of means to respond to this stress.

Here we will review the evidence for and mechanisms of host-mediated amino acid starvation and the bacterial response to amino acid depletion. We have chosen three intracellular pathogens to illustrate the diversity of bacterial responses. The first, *Chlamydia trachomatis,* is a natural auxotroph for many amino acids and responds to amino acid starvation primarily by differentiating to a viable but non-replicating form. The second, *Mycobacterium tuberculosis*, has evolved to become independent of host amino acid availability and constitutively

synthesizes its own amino acids. The third, *Legionella pneumophila*, is able to extract amino acids from an otherwise stingy host by exploiting host machinery—including amino acid transporters and the host proteasome—to make amino acids available in the pathogen's intracellular niche.

Host cells are able to starve intracellular pathogens of amino acids

The Evidence

The exact intracellular niches of bacterial pathogens are difficult to isolate; thus their metabolite contents are difficult to measure directly. Pathogen adaptation to the intracellular environment, however, suggests that host cells are able to limit amino acid availability. The most compelling fact is that amino acid auxotroph strains of many bacterial pathogens are often attenuated for intracellular growth and during host infection. For example, *Mycobacterium tuberculosis* requires the biosynthesis pathways for at least four amino acids to survive during a model infection of mice and within isolated mouse and human macrophages. Leucine, proline and lysine auxotrophs are attenuated *in vivo,* the latter by so much that they have been tested as live-attenuated vaccine candidates(Hondalus *et al.*, 2000, Smith *et al.*, 2001, Pavelka *et al.*, 2003b). Additionally, multiple groups have shown the requirement of several tryptophan biosynthesis enzymes for survival in mice and macrophages(Smith *et al.*, 2001, Parish, 2003). *Salmonella* auxotrophs for aromatic amino acids, histidine and methionine have also been shown to be attenuated *in vivo* (Hoiseth *et al.*, 1981, Fields *et al.*, 1986, O'Callaghan *et al.*, 1988).

Measuring the pathogen's transcriptional response to the intracellular niche paints the same picture of amino acid starvation. Since pathogens often regulate gene expression in response to environmental signals, expression profiles can serve as bioprobes of the conditions they encounter. Many intracellular pathogens upregulate amino acid biosynthesis genes during

infection(Chatterji *et al.*, 2001). This is true for bacterial pathogens as well as fungal pathogens like *Candida albicans(Rubin-Bejerano et al., 2003)*. Many fungal species encode a kinase, Gcn4, that senses exogenous amino acid availability. Upon infection, Gcn4 is activated and coordinates *Candida* amino acid biosynthesis(Tripathi *et al.*, 2002). Additionally, the stringent response, a conserved bacterial signaling pathway that senses amino acid starvation, among other stresses, is often required for intracellular growth(Magnusson *et al.*, 2005, Potrykus *et al.*, 2008). RelA senses amino acid starvation and synthesizes ppGpp, which induces necessary adaptive changes, including transcriptional upregulation of amino acid biosynthesis genes(Traxler *et al.*, 2008). In many intracellular pathogens, including M. tuberculosis and *Salmonella,* RelA is upregulated and specifically required for intracellular growth(Pizarro-Cerdá *et al.*, 2004, Song *et al.*, 2004, Primm *et al.*, 2000, Thompson *et al.*, 2006).

The Mechanisms

While there is much evidence that the host can create an amino acid depleted niche, the mechanisms for producing this starved niche are not well characterized. One major clue, however, comes from the autophagic response to intracellular pathogens. Autophagy is an important cellular process through which eukaryotic cells respond to a variety of stresses, including metabolite starvation and damaged organelles(Deretic, 2009, Deretic *et al.*, 2009, Levine, 2005). Bacterial invasion also induces an autophagic response, called xenophagy, wherein invading pathogens are delivered to autophagosomes to be broken down(Deretic, 2009, Deretic *et al.*, 2009, Levine, 2005). This process of

xenophagy is required for optimal suppression of the growth of multiple intracellular pathogens, including *Listeria*, *Shigella*, and *Mycobacterium* species(Suzuki *et al.*, 2007, Zhao *et al.*, 2008, Gutierrez *et al.*, 2004, Singh *et al.*, 2006, Singh *et al.*, 2010, Yuk *et al.*, 2009).

Nutritional signals play a key regulatory role in autophagy. In particular, autophagy is inhibited when the cytosol contains high concentrations of amino acids. Tattoli and colleagues found that intracellular amino acid levels must be low in order for xenophagy to be fully induced(Sancak *et al.*, 2010, Sancak *et al.*, 2008). They found that the activity of mTOR, a checkpoint kinase that links nutrient sensing to metabolic activity, decreased upon infection with the intracellular pathogens *Salmonella* and *Shigella* (Tattoli *et al.*, 2012b). This indicated a possible amino acid starvation, as mTOR activity decreases when amino acid levels are low. Moreover, cytosolic levels of isoleucine and leucine fell within an hour after infection(Tattoli *et al.*, 2012b). Interestingly, *Shigella*-induced amino acid starvation lasted at least four hours, while *Salmonella* induced a more transient one-hour starvation. The authors hypothesized that membrane damage might be responsible for amino acid starvation as the temporal pattern of such damage (transient for *Salmonella*; lasting for *Shigella*) matched the loss of amino acids. In fact, even aseptic membrane damage, induced by digitonin, could induce amino acid starvation. While this does not rule out alternative mechanisms, this study strongly supports the possibility that bacteria induce amino acid starvation and presents compelling evidence that membrane damage could play a major role(Tattoli *et al.*, 2012b, Tattoli *et al.*, 2012a). Interestingly,

Salmonella could reverse this amino acid starvation, which activated mTOR and inhibited autophagy, ultimately restoring bacterial growth. When the authors overrode the starvation-induced autophagy block, Salmonella growth was restored (Tattoli *et al.*, 2012b).

In addition to general amino acid depletion, host cells can also specifically deplete intracellular tryptophan during pathogen invasion(Hayashi *et al.*, 2001, Silva *et al.*, 2002). Tryptophan depletion as an anti-bacterial mechanism was first described in *Chlamydia*-infected cells(Beatty *et al.*, 1994, Byrne *et al.*, 1986, Murray *et al.*, 1989). The addition of interferon-gamma (IFN-γ) to infected cells almost completely abolishes intracellular chlamydial growth. Stunningly, simply adding tryptophan reverses the anti-chlamydial effect of IFN-γ(Beatty *et al.*, 1994, Byrne *et al.*, 1986, Leonhardt *et al.*, 2007). IFN-γ activates indoleamine-2,3-dioxygenase (IDO), the enzyme responsible for the first step in kynurenine synthesis from tryptophan(Zelante *et al.*, 2009, Leonhardt *et al.*, 2007). When strongly induced—as in the case of IFN-γ activation—IDO can deplete about 95% of intracellular tryptophan (Fujigaki, 2002). Thus, IDO is required for limiting the growth of tryptophan-requiring intracellular pathogens, such as some *Chlamydia* species, *Leishmania donovani*, *Toxoplasma gondii,* as well as lab-generated strains tryptophan-auxotrophic bacteria(Daubener *et al.*, 2001, Fujigaki, 2002, Leonhardt *et al.*, 2007, Ibana *et al.*, 2011b, Ibana *et al.*, 2011a).

It is worth noting, however, that the effect of IDO on pathogen growth is multi-faceted(Medzhitov *et al.*, 2011). The first appreciated role for IDO in pathogen defense was intracellular tryptophan depletion. Recently, the role of its

enzymatic products, kynurenines, has also been characterized. Kynurenines are potent negative regulators of inflammation and T cell activity(Munn *et al.*, 2005, Zelante *et al.*, 2009, Medzhitov *et al.*, 2011, Favre *et al.*, 2010). So while the role of IDO on intracellular tryptophan is clear, its pleiotropic effects in pathogen defense have been more difficult to pin down(Blumenthal *et al.*, 2012). Nevertheless, IDO represents a mechanism by which an infected host can drastically reduce the amounts of an often-essential amino acid available to pathogen.

Pathogen Responses to Host AA Starvation

Many pathogens are able to synthesize the full set of amino acids. It would be reasonable to think, given how common amino acid depleted environments are, that amino acid biosynthetic pathways would be a requirement for being an intracellular human pathogen. However, many known successful human pathogens are auxotrophic for some amino acids. How can this be? Some bacteria are simply able to escape by altering their metabolic requirements. For example, they can differentiate into non-replicating, less metabolically demanding cells. Other organisms are able to avoid amino acid starvation altogether by manipulating host amino acid uptake and production systems to increase available amino acids in the pathogen's intracellular niche. We will review three model pathogens that fit these three molds: 1.) *Chlamydia*, which employs growth arrest and differentiation; 2.) *Mycobacterium tuberculosis*, which exemplifies amino acid self-sufficiency; and 3.) *Legionella pneumophila*, which exploits host machinery to extract amino acids from the host cell.

Chlamydia: Hiding out and holding out for a better day—growth arrest and differentiation

Chlamydia trachomatis causes a variety of human diseases, including the leading cause of infectious blindness and genital tract infections that can have long-term severe consequences for infected individuals. All members of the *Chlamydia* family are intracellular pathogens that share two distinct developmental stages. Elementary bodies (EB) initiate the intracellular infection through endocytic uptake. In optimal growth conditions, the metabolically inert EBs then differentiate into metabolically active reticulate bodies (RB), which then replicate and grow within the bacterial vacuole. Eventually RBs develop back into EBs, which upon release, go on to infect other cells.

When stimulated with IFN-γ, the host cell imposes a much greater stress upon the bacteria. *Chlamydia* responds to this stress by morphing into a third, aberrant form(Leonhardt *et al.*, 2007, Beatty *et al.*, 1994). The aberrant form cells are similar to reticulate bodies in morphology, but are unable to divide or differentiate back to EBs. In some studies, adding tryptophan restores growth of these aberrant RBs, demonstrating that persistence is driven by tryptophan starvation(Byrne *et al.*, 1986, Beatty *et al.*, 1994, Leonhardt *et al.*, 2007, Ibana *et al.*, 2011a). Differentiation into this aberrant RB form allows the pathogen to hide from its tryptophan-depleted environment until tryptophan availability resumes (Figure 1.1). Isoleucine also seems to be limited during infection, and *Chlamydia* growth in one intracellular infection model could be restored upon the addition of isoleucine (Hatch, 1975). Interestingly, *Chlamydia* does not encode the classical

bacterial stringent response, which enables many bacteria to sense amino acid starvation and enter into persistent states(Ouellette *et al.*, 2006). Thus, *Chlamydia* must utilize a unique mechanism to sense tryptophan depletion and differentiate into these aberrant forms.

Early investigations on the role of tryptophan and IDO in driving aberrant form differentiation yielded mixed results. While some groups found tryptophan supplementation alone could reverse the effects of IFN-γ, others found no effect(Murray *et al.*, 1989). When the *Chlamydia trachomatis* genome was sequenced—and especially when multiple clinical strains were sequenced—a major clue emerged. As expected, *Chlamydia* lacks the full suite of tryptophan biosynthetic enzymes, and thus cannot synthesize tryptophan from chorismate or most other biosynthetic precursors(Stephens, 1998). It does, however, encode tryptophan synthase (TrpAB), a more limited enzyme that allows the organism to synthesize tryptophan from indole. In addition *Chlamydia* encodes TrpR, a tryptophan-dependent aporepressor, which enables the pathogen to react specifically to environmental tryptophan availability(Stephens, 1998, Wood *et al.*, 2003, Belland *et al.*, 2003). Interestingly, *C. trachomatis* strains isolated from ocular infections and synovial tissue contain frameshift mutations in *trpAB*, while strains from genital infections do not(Caldwell *et al.*, 2003, Fehlner-Gardiner *et al.*, 2002, Gerard *et al.*, 2010). Within strains isolated from sexually transmitted infections, there are three *trpAB* variants; while the effect of each *trpAB* variant on tryptophan synthesis is unknown, identity of the *trpAB* allele correlates perfectly with IFN-γ susceptibility.*in vitro*(Morrison, 2000

, Caldwell *et al.*, 2003, Fehlner-Gardiner *et al.*, 2002, McClarty *et al.*, 2007). So while all all *C. trachomatis* contain the *trpAB* gene, not all strains are able to make tryptophan, and the ability to make tryptophan appears to track with STI virulence.

Is the *C. trachomatis* response to tryptophan starvation differentiation to a non-growing cell type or regulated tryptophan biosynthesis? It appears that tryptophan biosynthesis is possible, but under a limited number of circumstances. First, the strain has to encode a functional enzyme. Second, it has to have a source of indole, which the host does not make. Some investigators have noticed that many functional TrpAB-containing strains were isolated from patients with bacterial vaginosis, and thus hypothesize that the vaginal flora might be a source of indole(McClarty *et al.*, 2007). So some *C. trachomatis* strains can synthesize tryptophan in certain environments, increasing their resistance to IFN-γ mediated immunity(Belland *et al.*, 2003). Where tryptophan biosynthesis is not available, *Chlamydia* use their ability to enter into a persistent form that enables it to survive when all sources of tryptophan are lost.

***M. tuberculosis*: Energy independence—making its own amino acids in an unreliable environment**

Whereas *Chlamydia* tryptophan biosynthesis is limited in scope, *Mycobacterium tuberculosis* seems to leave nothing up to chance. It contains the entire biosynthetic toolset for all 20 amino acids(Cole *et al.*, 1998). During infection of both mice and macrophages, it appears that *M. tuberculosis* constitutively expresses amino acid biosynthesis genes(Schnappinger *et al.*,

2003, Rohde *et al.*, 2012, Talaat *et al.*, 2004, Talaat *et al.*, 2007). Even when grown *in vitro*, where tryptophan is freely available, the tryptophan biosynthetic locus continues to be expressed [unpublished data, Y. Zhang and E. Rubin]. Thus, *M. tuberculosis* seems to choose to independently synthesize amino acids regardless of environmental availability (Figure 1). *M. tuberculosis* does the same with energetic sources as well—unlike *E. coli*, *M. tuberculosis* will continue to metabolize energy-poor carbon sources even in the presence of more energy efficient carbon sources(de Carvalho *et al.*, 2010).

The *M. tuberculosis* intracellular niche is similar to *Chlamydia*'s. *M. tuberculosis* is taken up by phagocytosis and enters into a vacuole that escapes lysosome fusion. As with *Chlamydia*, there is strong evidence that IFN-γ activation of infected macrophages plays a major role in controlling *M. tuberculosis* growth(Flynn *et al.*, 1993, Fabri *et al.*, 2011). Unlike *Chlamydia*, *M. tuberculosis* is always able to make tryptophan, which is essential for its survival of IDO activation by IFN-γ. Its ability to make other amino acids is also crucial for virulence, as lysine, proline and leucine auxotrophs also fail to grow normally in macrophages and are severely attenuated in mice(Hondalus *et al.*, 2000, Smith *et al.*, 2001, Parish, 2003, Pavelka *et al.*, 2003a). Perhaps because amino acid biosynthesis is constitutive in *M. tuberculosis*, amino acid biosynthesis has not been seen as a virulence factor. However, these processes are clearly essential and, in fact, could serve as the basis for drug design.

If amino acid production is constitutive, is there any role for amino acid sensing during infection? *M. tuberculosis* does have a stringent response, and it

encodes a RelA homologue, which in *E. coli* and other organisms responds to amino acid starvation(Primm *et al.*, 2000). RelA knockouts are attenuated for growth in mice and guinea pigs(Primm *et al.*, 2000, Klinkenberg *et al.*, 2010, Dahl *et al.*, 2003). It is likely that RelA can respond to amino acid levels—this is true in a related species, *Mycobacterium smegmatis*—but neither RelA sensing of amino acid starvation nor RelA-dependent coordination of amino acid synthesis is well documented in M. tuberculosis(Dahl *et al.*, 2005). The stringent response signals through ppGpp, which in *E. coli* requires DksA to be fully active(Magnusson *et al.*, 2005, Paul *et al.*, 2004). *M. tuberculosis* encodes CarD, an alternative DksA, which is also required for the stringent response to starvation, and mutants are slightly attenuated in mice(Connolly *et al.*, 2009). However, CarD is essential in rich media, suggesting a role outside of the stringent response that might explain its requirement for growth *in vivo*(Connolly *et al.*, 2009, Stallings *et al.*, 2009). Notably, RelA mutants are far less attenuated than amino acid auxotrophs(Primm *et al.*, 2000). It is most likely that the stringent response in *M. tuberculosis* is primarily responsible for sensing and responding to other forms of starvation or immune-mediated stress. Thus, *M. tuberculosis* is an example of a pathogen that has decoupled amino acid metabolism from host availability. It persists in making its own amino acids, making it resistant to host mechanisms of amino acid starvation.

Legionella: Host exploitation—extracting amino acids from a stingy host

Not all bacteria are as energy independent as *M. tuberculosis*. The host, after all, is a tremendous resource for essential nutrients if the pathogen can

manage access(Rohmer *et al.*, 2011). In this respect, *Legionella pneumophila* has been quite successful. Like *M. tuberculosis* and *Chlamydia, L. pneumophila* replicates in an intracellular vacuole, the *Legionella*-containing vacuole (LCV). The LCV avoids lysosomal fusion and further develops by recruiting mitochondria and rough ER-derived vesicles. *L. pneumophila* translocates ~300 effectors into the host cell cytoplasm via its type IV secretion system to specifically refine its intracellular niche(Al-Quadan *et al.*, 2012).

Unlike its other intracellular pathogen cousins, the evidence seems to suggest that *L. pneumophila* does not face an amino acid starved milieu. *L. pneumophila* does employ the stringent response, which is necessary for survival during infection. In fact, the stringent response in *L. pneumophila* drives the expression of most known virulence factors(Hammer *et al.*, 1999, Molofsky *et al.*, 2004). However, in macrophages, its stringent response machinery does not rely on RelA, suggesting that sensing amino acid starvation is unimportant during infection(Dalebroux *et al.*, 2009, Dalebroux *et al.*, 2010). Instead *L. pneumophila* responds to fatty acid starvation through another stringent response sensor, SpoT, and uses this as a signal to turn on its intracellular pathogenesis program(Dalebroux *et al.*, 2009). Amino acid catabolism is necessary for growth *in vivo* and some evidence points to serine as a key intracellular carbon source for *L. pneumophila*(Eylert *et al.*, 2010). Finally, a bacterial transporter, PhtA, is used by *L. pneumophila* as a high-affinity threonine transporter and is necessary for growth in cells, showing that at least one amino acid is acquired from the host(Sauer *et al.*, 2005).

The LCV, then, seems to be an amino acid rich environment designed for a hungry bacterium. How then, does *L. pneumophila* manipulate host cells, normally stingy with their amino acids, to serve up amino acids into the LCV? After all, it is likely that amino acid starvation programs are triggered in host cells infected with *L. pneumophila*. The bacterium has at least two characterized mechanisms of reversing this starvation.

First, *L. pneumophila* upregulates a host neutral amino acid transporter, SLC1A5, during macrophage infection (Figure 1.1) Knockdown of the transporter does not affect host cell viability, but limits *Legionella* growth(Wieland *et al.*, 2005). Even when cells were grown in amino acid rich media, an 80% decrease in SLC1A5 expression resulted in ~1000 fold decreased *L. pneumophila* growth(Wieland *et al.*, 2005). The same transporter is specifically upregulated during infection with another intracellular pathogen, *Francisella*, and is also required for bacterial replication in THP-1 cells(Barel *et al.*, 2012). It is unclear whether SLC1A5 exists on the LCV membrane, the plasma membrane, or both, but it is essential in providing amino acids to *L. pneumophila.*

Second, a recent study of host-pathogen interactions at the LCV membrane revealed that *L. pneumophila* utilizes the host proteasome to generate free amino acids for bacterial growth(Price *et al.*, 2011). The authors noted that the LCV is decorated with Lys48-linked polyubiquitinated proteins anchored to the bacterial virulence factor AnkB(Price *et al.*, 2011). These polyubiquitinated proteins recruit the host proteasome, generating small peptides at the LCV membrane. Inhibition of both the proteasome and amino acid-

releases peptidases results arrests *L. pneumophila* growth, a phenomenon that is reversed by the addition of excess free amino acids(Price *et al.*, 2011). Using a standard transcriptional reporter of amino acid starvation (green fluorescent protein fused to the *flaA* promoter), the authors found that while wildtype *L. pneumophila* did not experience amino acid starvation during infection—confirming previous stringent response studies—AnkB null mutants did. AnkB growth was not further suppressed by proteasome inhibition, and was rescued by amino acid supplementation. Thus, by recruiting proteins and targeting them for proteasomal degradation, AnkB provides a ready source of amino acids(Price *et al.*, 2011). *L. pneumophila* actively molds the LCV into a growth-supporting niche, subverting host amino acid starvation machinery and exploiting host AA-acquisition mechanisms(Al-Quadan *et al.*, 2012).

Concluding Remarks

Many of us take a very pathogen-centric view of host defense. However, the vast majority of bacteria encountered by humans are incapable of causing infection. "Non-specific" defenses, such as nutrient deprivation, play a large role in eliminating organisms that are not specifically adapted to a pathogenic niche. Pathogens, therefore, have evolved specific mechanisms to subvert these defenses and to take advantage of the nutrient rich human host. Amino acids, necessary for protein synthesis and, sometimes, as a carbon source, are often depleted in the intracellular environment as a means of starving the pathogen. Pathogen responses range from growth suppression to manipulating host pathways to reverse amino acid starvation.

Understanding the interplay between host and pathogen could prove to be useful in designing new therapies(Rasko *et al.*, 2010, Escaich, 2008). The three bacteria discussed in this review, *C. trachomatis*, *L. pneumophila*, and *M. tuberculosis* are major human pathogens that, for various reasons, have proven to be difficult to control. All three represent organisms for which new antibacterial therapies are needed. By knowing how our immune system tries to kill these bacteria as well as the bacterial evasion strategies, we could target these processes to synergize with the immune system to kill invading pathogens. For example, blocking tryptophan biosynthesis, could increase IFN-γ- mediated killing of *M. tuberculosis*. Amino acid biosynthesis genes could be reasonable drug targets. It is already targeted by many herbicides, and a small molecule used to treat mouse *Pseudomonas aureginosa* infection was shown to have activity

against bacterial tryptophan biosynthesis(Lesic *et al.*, 2007, Epelbaum *et al.*, 1996). Furthermore, host tryptophan needs are supplied through the diet, suggesting that, barring off-target effects, small molecules targeting tryptophan synthesis will not be toxic to the host. Alternatively, the susceptibility of *Legionella* to the inhibition of the proteasome suggests that host-directed therapies could alter that balance in virulence(Price *et al.*, 2011, Al-Quadan *et al.*, 2012). Thus, altering both bacterial and host responses critical for amino acid starvation could provide new avenues for the development of therapeutics.

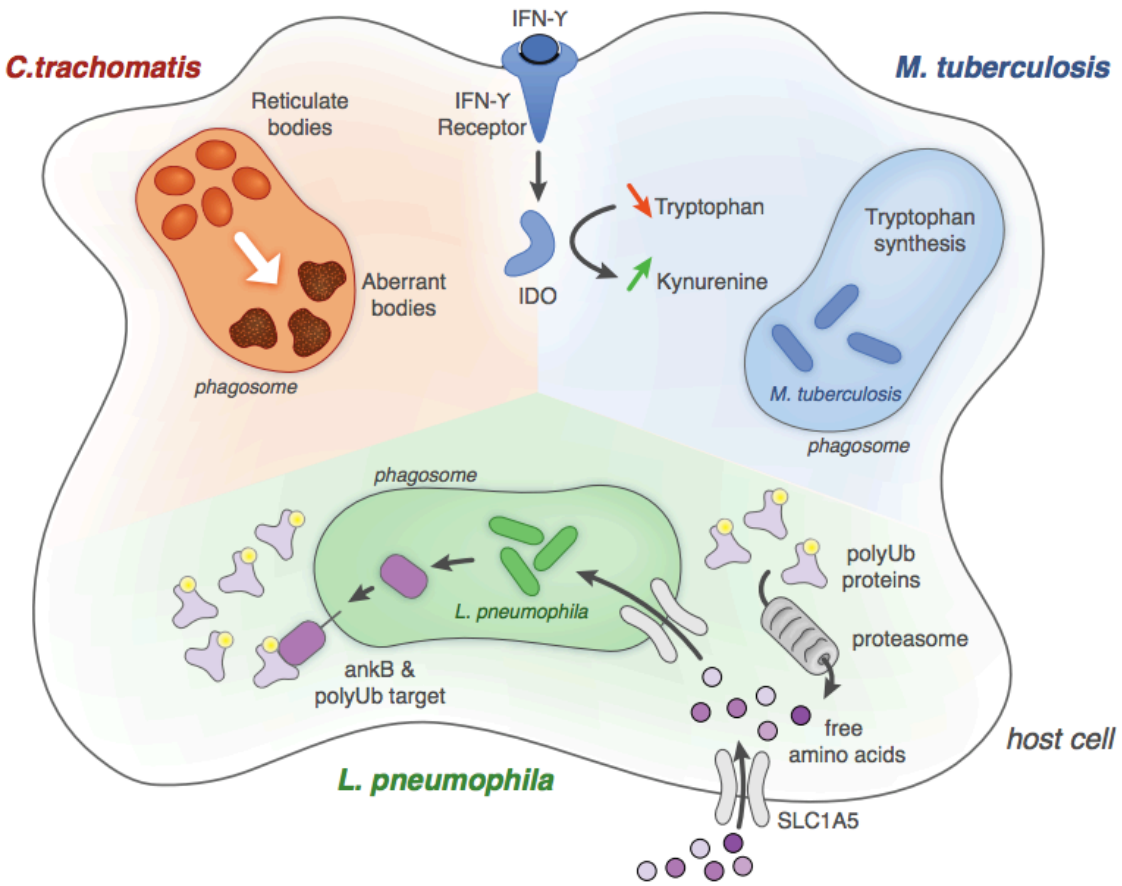


Figure 1.1: Three bacterial strategies for evading amino acid starvation
Chlamydia trachomatis responds to amino acid starvation by differentiating to aberrant reticulate bodies, viable but metabolically inactive forms that can differentiate back to active forms when amino acid become available again. While *Mycobacterium tuberculosis* faces amino acid starvation during infection, its constitutive expression of amino acid synthesis makes it generally resistant to host-driven starvation. *Legionella pneumophila* exploits host proteins to override starvation and deliver free amino acids into its intracellular niche. AnkB, a secreted virulence factor, drives the polyubiquitination of LCV membrane proteins. The host proteasome is then recruited to proteolyze ubiquinated targets, producing free amino acids that are transported into the LCV by the upregulated host transporter, SLC1A5.

Section 1.3. Bibliography (I)

Al-Quadan, T., Price, C.T. and Abu Kwaik, Y. (2012). Exploitation of evolutionarily conserved amoeba and mammalian processes by Legionella. *Trends in microbiology* **20,** 299-306.

Barel, M., Meibom, K., Dubail, I., Botella, J. and Charbit, A. (2012). Francisella tularensis regulates the expression of the amino acid transporter SLC1A5 in infected THP-1 human monocytes. *Cellular microbiology*.

Beatty, W.L., Belanger, T.A., Desai, A.A., Morrison, R.P. and Byrne, G.I. (1994). Tryptophan depletion as a mechanism of gamma interferon-mediated chlamydial persistence. *Infection and immunity* **62,** 3705-3711.

Belland, R.J., Zhong, G., Crane, D.D., Hogan, D., Sturdevant, D., Sharma, J.*, et al.* (2003). Genomic transcriptional profiling of the developmental cycle of Chlamydia trachomatis. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 8478-8483.

Blumenthal, A., Nagalingam, G., Huch, J.H., Walker, L., Guillemin, G.J., Smythe, G.A.*, et al.* (2012). M. tuberculosis induces potent activation of IDO-1, but this is not essential for the immunological control of infection. *PloS one* **7,** e37314.

Byrne, G.I., Lehmann, L.K. and Landry, G.J. (1986). Induction of tryptophan catabolism is the mechanism for gamma-interferon-mediated inhibition of intracellular Chlamydia psittaci replication in T24 cells. *Infection and immunity* **53,** 347-351.

Caldwell, H.D., Wood, H., Crane, D., Bailey, R., Jones, R.B., Mabey, D.*, et al.* (2003). Polymorphisms in Chlamydia trachomatis tryptophan synthase genes differentiate between genital and ocular isolates. *The Journal of clinical investigation* **111,** 1757-1769.

Chatterji, D. and Ojha, A. (2001). Revisiting the stringent response, ppGpp and starvation signaling. *Curr Opin Microbiol* **4,** 160-165.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D.*, et al.* (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393,** 537-544.

Connolly, L.E. and Cox, J.S. (2009). CarD tricks and magic spots: mechanisms of stringent control in mycobacteria. *Cell Host and Microbe* **6,** 1-2.

Dahl, J.L., Arora, K., Boshoff, H.I., Whiteford, D.C., Pacheco, S.A., Walsh, O.J., *et al.* (2005). The relA homolog of Mycobacterium smegmatis affects cell appearance, viability, and gene expression. *J Bacteriol* **187,** 2439-2447.

Dahl, J.L., Kraus, C.N., Boshoff, H.I., Doan, B., Foley, K., Avarbock, D., *et al.* (2003). The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of Mycobacterium tuberculosis in mice. *Proc Natl Acad Sci U S A* **100,** 10026-10031.

Dalebroux, Z.D., Edwards, R.L. and Swanson, M.S. (2009). SpoT governs Legionella pneumophila differentiation in host macrophages. *Molecular microbiology* **71,** 640-658.

Dalebroux, Z.D., Yagi, B.F., Sahr, T., Buchrieser, C. and Swanson, M.S. (2010). Distinct roles of ppGpp and DksA in Legionella pneumophila differentiation. *Molecular microbiology* **76,** 200-219.

Daubener, W., Spors, B., Hucke, C., Adam, R., Stins, M., Kim, K.S. and Schroten, H. (2001). Restriction of Toxoplasma gondii growth in human brain microvascular endothelial cells by activation of indoleamine 2,3-dioxygenase. *Infect Immun* **69,** 6527-6531.

de Carvalho, L.P., Zhao, H., Dickinson, C.E., Arango, N.M., Lima, C.D., Fischer, S.M., *et al.* (2010). Activity-based metabolomic profiling of enzymatic function: identification of Rv1248c as a mycobacterial 2-hydroxy-3-oxoadipate synthase. *Chemistry & biology* **17,** 323-332.

Deretic, V. (2009). Strange bedfellows expose ancient secrets of autophagy in immunity. *Immunity* **30,** 479-481.

Deretic, V. and Levine, B. (2009). Autophagy, immunity, and microbial adaptations. *Cell host & microbe* **5,** 527-549.

Eisenreich, W., Dandekar, T., Heesemann, J. and Goebel, W. (2010). Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. *Nature reviews. Microbiology* **8,** 401-412.

Epelbaum, S., Chipman, D. and Barak, Z.e. (1996). Metabolic Effects of Inhibitors of Two Enzymes of the Branched-Chain Amino Acid Pathway in Salmonella typhimurium. *Journal of Bacteriology* **178,** 1187-1196.

Escaich, S. (2008). Antivirulence as a new antibacterial approach for chemotherapy. *Current opinion in chemical biology* **12,** 400-408.

Eylert, E., Herrmann, V., Jules, M., Gillmaier, N., Lautner, M., Buchrieser, C., *et al.* (2010). Isotopologue profiling of Legionella pneumophila: role of serine

and glucose as carbon substrates. *Journal of Biological Chemistry* **285,** 22232-22243.

Fabri, M., Stenger, S., Shin, D.M., Yuk, J.M., Liu, P.T., Realegeno, S.*, et al.* (2011). Vitamin D is required for IFN-gamma-mediated antimicrobial activity of human macrophages. *Sci Transl Med* **3,** 104ra102.

Favre, D., Mold, J., Hunt, P.W., Kanwar, B., Loke, P.a.n., Seu, L.*, et al.* (2010). Tryptophan catabolism by indoleamine 2,3-dioxygenase 1 alters the balance of TH17 to regulatory T cells in HIV disease. *Science Translational Medicine* **2,** 32ra36.

Fehlner-Gardiner, C., Roshick, C., Carlson, J.H., Hughes, S., Belland, R.J., Caldwell, H.D. and McClarty, G. (2002). Molecular basis defining human Chlamydia trachomatis tissue tropism. A possible role for tryptophan synthase. *The Journal of biological chemistry* **277,** 26893-26903.

Fields, P.I., Swanson, R.V., Haidaris, C.G. and Heffron, F. (1986). Mutants of Salmonella typhimurium that cannot survive within the macrophage are avirulent. *Proceedings of the National Academy of Sciences of the United States of America* **83,** 5189-5193.

Flynn, J.L., J, C., Triebold, K.J., Dalton, D.K., Stewart, T.A. and Bloom, B.R. (1993). An Essential Role for Interferon gamma in Resistance to Mycobacterium tuberculosis Infection. *Journal of Expermental Medicine* **178**.

Fujigaki, S. (2002). L-Tryptophan-L-Kynurenine Pathway Metabolism Accelerated by Toxoplasmagondii Infection Is Abolished in Gamma Interferon-Gene-Deficient Mice: Cross-Regulation between Inducible Nitric Oxide Synthase and Indoleamine-2,3-Dioxygenase. *Infection and immunity* **70,** 779-786.

Gerard, H.C., Stanich, J.A., Whittum-Hudson, J.A., Schumacher, H.R., Carter, J.D. and Hudson, A.P. (2010). Patients with Chlamydia-associated arthritis have ocular (trachoma), not genital, serovars of C. trachomatis in synovial tissue. *Microb Pathog* **48,** 62-68.

Gutierrez, M.G., Master, S.S., Singh, S.B., Taylor, G.A., Colombo, M.I. and Deretic, V. (2004). Autophagy is a defense mechanism inhibiting BCG and Mycobacterium tuberculosis survival in infected macrophages. *Cell* **119,** 753-766.

Hammer, B.K. and Swanson, M.S. (1999). Co-ordination of Legionella pneumophila virulence with entry into stationary phase by ppGpp. *Molecular microbiology* **33,** 721-731.

Hatch, T.P. (1975). Competition Between Chlamydia psitacci and L Cells for Host Isoleucine Pools: a Limiting Factor in Chlamydial Multiplication. *Infection and immunity* **12,** 211-220.

Hayashi, T., Rao, S.P., Takabayashi, K., Van Uden, J.H., Kornbluth, R.S., Baird, S.M.*, et al.* (2001). Enhancement of innate immunity against Mycobacterium avium infection by immunostimulatory DNA is mediated by indoleamine 2,3-dioxygenase. *Infect Immun* **69,** 6156-6164.

Hoiseth, S.K. and Stocker, B.A. (1981). Aromatic-dependent Salmonella typhimurium are non-virulent and effective as live vaccines. *Nature* **291,** 238-239.

Hondalus, M.K., Bardarov, S., Russell, R., Chan, J., Jacobs, W.R. and Bloom, B.R. (2000). Attenuation of and protection induced by a leucine auxotroph of Mycobacterium tuberculosis. *Infection and immunity* **68,** 2888-2898.

Hood, M.I. and Skaar, E.P. (2012). Nutritional immunity: transition metals at the pathogen-host interface. *Nature reviews. Microbiology* **10,** 525-537.

Ibana, J.A., Belland, R.J., Zea, A.H., Schust, D.J., Nagamatsu, T., AbdelRahman, Y.M.*, et al.* (2011a). Inhibition of indoleamine 2,3-dioxygenase activity by levo-1-methyl tryptophan blocks gamma interferon-induced Chlamydia trachomatis persistence in human epithelial cells. *Infect Immun* **79,** 4425-4437.

Ibana, J.A., Belland, R.J., Zea, A.H., Schust, D.J., Nagamatsu, T., AbdelRahman, Y.M.*, et al.* (2011b). Inhibition of indoleamine 2,3-dioxygenase activity by levo-1-methyl tryptophan blocks gamma interferon-induced Chlamydia trachomatis persistence in human epithelial cells. *Infection and immunity* **79,** 4425-4437.

Klinkenberg, L.G., Lee, J.H., Bishai, W.R. and Karakousis, P.C. (2010). The stringent response is required for full virulence of Mycobacterium tuberculosis in guinea pigs. *The Journal of Infectious Diseases* **202,** 1397-1404.

Leonhardt, R.M., Lee, S.-J., Kavathas, P.B. and Cresswell, P. (2007). Severe tryptophan starvation blocks onset of conventional persistence and reduces reactivation of Chlamydia trachomatis. *Infection and immunity* **75,** 5105-5117.

Lesic, B., Lepine, F., Deziel, E., Zhang, J., Zhang, Q., Padfield, K.*, et al.* (2007). Inhibitors of pathogen intercellular signals as selective anti-infective compounds. *PLoS Pathog* **3,** 1229-1239.

Levine, B. (2005). Eating oneself and uninvited guests: autophagy-related pathways in cellular defense. *Cell* **120,** 159-162.

Magnusson, L.U., Farewell, A. and Nyström, T. (2005). ppGpp: a global regulator in Escherichia coli. *Trends in microbiology* **13,** 236-242.

McClarty, G., Caldwell, H.D. and Nelson, D.E. (2007). Chlamydial interferon gamma immune evasion influences infection tropism. *Current opinion in microbiology* **10,** 47-51.

Medzhitov, R., Shevach, E.M., Trinchieri, G., Mellor, A.L., Munn, D.H., Gordon, S.*, et al.* (2011). Highlights of 10 years of immunology in Nature Reviews Immunology. *Nature reviews Immunology* **11,** 693-702.

Molofsky, A.B. and Swanson, M.S. (2004). Differentiate to thrive: lessons from the Legionella pneumophila life cycle. *Molecular microbiology* **53,** 29-40.

Morrison, R.P. (2000

). Differential Sensitivities of Chlamydia trachomatis Strains to Inhibitory Effects of Gamma Interferon. *Infection and immunity* **68,** 6038.

Munn, D.H., Sharma, M.D., Baban, B., Harding, H.P., Zhang, Y., Ron, D. and Mellor, A.L. (2005). GCN2 Kinase in T Cells Mediates Proliferative Arrest and Anergy Induction in Response to Indoleamine 2,3-Dioxygenase. *Immunity* **22,** 633-642.

Murray, H.W., Szuro-Sudol, A., Wellner, D., Oca, M.J., Granger, A.M., Libby, D.M.*, et al.* (1989). Role of tryptophan degradation in respiratory burst-independent antimicrobial activity of gamma interferon-stimulated human macrophages. *Infection and immunity* **57,** 845-849.

O'Callaghan, D., Maskell, D., Liew, F.Y., Easmon, C.S. and Dougan, G. (1988). Characterization of aromatic- and purine-dependent Salmonella typhimurium: attention, persistence, and ability to induce protective immunity in BALB/c mice. *Infection and immunity* **56,** 419-423.

Ouellette, S.P., Hatch, T.P., AbdelRahman, Y.M., Rose, L.A., Belland, R.J. and Byrne, G.I. (2006). Global transcriptional upregulation in the absence of increased translation in Chlamydia during IFNgamma-mediated host cell tryptophan starvation. *Molecular microbiology* **62,** 1387-1401.

Parish, T. (2003). Starvation survival response of Mycobacterium tuberculosis. *Journal of Bacteriology* **185,** 6702-6706.

Paul, B.J., Barker, M.M., Ross, W., Schneider, D.A., Webb, C., Foster, J.W. and Gourse, R.L. (2004). DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* **118,** 311-322.

Pavelka, M.S., Chen, B., Kelley, C.L., Collins, F.M. and Jacobs Jr, W.R. (2003a). Vaccine efficacy of a lysine auxotroph of Mycobacterium tuberculosis. *Infection and immunity* **71,** 4190-4192.

Pavelka, M.S., Chen, B., Kelley, C.L., Collins, F.M. and Jacobs, W.R. (2003b). Vaccine Efficacy of a Lysine Auxotroph of Mycobacterium tuberculosis. *Infection and immunity* **71,** 4190-4192.

Pizarro-Cerdá, J. and Tedin, K. (2004). The bacterial signal molecule, ppGpp, regulates Salmonella virulence gene expression. *Molecular microbiology* **52,** 1827-1844.

Potrykus, K. and Cashel, M. (2008). (p)ppGpp: still magical? *Annual review of microbiology* **62,** 35-51.

Price, C.T.D., Al-Quadan, T., Santic, M., Rosenshine, I. and Abu Kwaik, Y. (2011). Host proteasomal degradation generates amino acids essential for intracellular bacterial growth. *Science (New York, NY)* **334,** 1553-1557.

Primm, T.P., Andersen, S.J., Mizrahi, V., Avarbock, D., Rubin, H. and Barry, C.E. (2000). The stringent response of Mycobacterium tuberculosis is required for long-term survival. *Journal of Bacteriology* **182,** 4889-4898.

Rasko, D.A. and Sperandio, V. (2010). Anti-virulence strategies to combat bacteria-mediated disease. *Nature reviews. Drug discovery* **9,** 117-128.

Rohde, K.H., Veiga, D.F.T., Caldwell, S., Balázsi, G. and Russell, D.G. (2012). Linking the Transcriptional Profiles and the Physiological States of Mycobacterium tuberculosis during an Extended Intracellular Infection. *PLoS Pathogens* **8,** e1002769.

Rohmer, L., Hocquet, D. and Miller, S.I. (2011). Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends Microbiol* **19,** 341-348.

Rubin-Bejerano, I., Fraser, I., Grisafi, P. and Fink, G.R. (2003). Phagocytosis by neutrophils induces an amino acid deprivation response in Saccharomyces cerevisiae and Candida albicans. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 11007-11012.

Sancak, Y., Bar-Peled, L., Zoncu, R., Markhard, A.L., Nada, S. and Sabatini, D.M. (2010). Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. *Cell* **141,** 290-303.

Sancak, Y., Peterson, T.R., Shaul, Y.D., Lindquist, R.A., Thoreen, C.C., Bar-Peled, L. and Sabatini, D.M. (2008). The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. *Science* **320,** 1496-1501.

Sauer, J.-D., Bachman, M.A. and Swanson, M.S. (2005). The phagosomal transporter A couples threonine acquisition to differentiation and replication of Legionella pneumophila in macrophages. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 9924-9929.

Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M.*, et al.* (2003). Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment. *The Journal of experimental medicine* **198,** 693-704.

Silva, N.M., Rodrigues, C.V., Santoro, M.M., Reis, L.F.L., Alvarez-Leite, J.I. and Gazzinelli, R.T. (2002). Expression of indoleamine 2,3-dioxygenase, tryptophan degradation, and kynurenine formation during in vivo infection with Toxoplasma gondii: induction by endogenous gamma interferon and requirement of interferon regulatory factor 1. *Infection and immunity* **70,** 859-868.

Singh, S.B., Davis, A.S., Taylor, G.A. and Deretic, V. (2006). Human IRGM induces autophagy to eliminate intracellular mycobacteria. *Science (New York, NY)* **313,** 1438-1441.

Singh, S.B., Ornatowski, W., Vergne, I., Naylor, J., Delgado, M., Roberts, E.*, et al.* (2010). Human IRGM regulates autophagy and cell-autonomous immunity functions through mitochondria. *Nature Cell Biology* **12,** 1154-1165.

Skaar, E.P. (2010). The Battle for Iron between Bacterial Pathogens and Their Vertebrate Hosts. *PLoS Pathogens*.

Smith, D.A., Parish, T., Stoker, N.G. and Bancroft, G.J. (2001). Characterization of auxotrophic mutants of Mycobacterium tuberculosis and their potential as vaccine candidates. *Infection and immunity* **69,** 1142-1150.

Song, M., Kim, H.-J., Kim, E.Y., Shin, M., Lee, H.C., Hong, Y.*, et al.* (2004). ppGpp-dependent stationary phase induction of genes on Salmonella pathogenicity island 1. *The Journal of biological chemistry* **279,** 34183-34190.

Stallings, C.L., Stephanou, N.C., Chu, L., Hochschild, A., Nickels, B.E. and Glickman, M.S. (2009). CarD is an essential regulator of rRNA transcription required for Mycobacterium tuberculosis persistence. *Cell* **138,** 146-159.

Stephens, R.S. (1998). Genome Sequence of an Obligate Intracellular Pathogen of Humans: Chlamydia trachomatis. *Science* **282,** 754-759.

Suzuki, T., Franchi, L., Toma, C., Ashida, H., Ogawa, M., Yoshikawa, Y.*, et al.* (2007). Differential regulation of caspase-1 activation, pyroptosis, and autophagy via Ipaf and ASC in Shigella-infected macrophages. *PLoS Pathog* **3,** e111.

Talaat, A.M., Lyons, R., Howard, S.T. and Johnston, S.A. (2004). The temporal expression profile of Mycobacterium tuberculosis infection in mice. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 4602-4607.

Talaat, A.M., Ward, S.K., Wu, C.-W., Rondon, E., Tavano, C., Bannantine, J.P.*, et al.* (2007). Mycobacterial bacilli are metabolically active during chronic tuberculosis in murine lungs: insights from genome-wide transcriptional profiling. *Journal of Bacteriology* **189,** 4265-4274.

Tattoli, I., Sorbara, M.T., Philpott, D.J. and Girardin, S.E. (2012a). Bacterial autophagy: The trigger, the target and the timing. *Autophagy* **8**.

Tattoli, I., Sorbara, M.T., Vuckovic, D., Ling, A., Soares, F., Carneiro, L.A.M.*, et al.* (2012b). Amino Acid Starvation Induced by Invasive Bacterial Pathogens Triggers an Innate Host Defense Program. *Cell Host and Microbe* **11,** 563-575.

Thompson, A., Rolfe, M.D., Lucchini, S., Schwerk, P., Hinton, J.C.D. and Tedin, K. (2006). The bacterial signal molecule, ppGpp, mediates the environmental regulation of both the invasion and intracellular virulence gene programs of Salmonella. *The Journal of biological chemistry* **281,** 30112-30121.

Traxler, M.F., Summers, S.M., Nguyen, H.T., Zacharia, V.M., Hightower, G.A., Smith, J.T. and Conway, T. (2008). The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli. *Mol Microbiol* **68,** 1128-1148.

Tripathi, G., Wiltshire, C., Macaskill, S., Tournu, H., Budge, S. and Brown, A.J.P. (2002). Gcn4 co-ordinates morphogenetic and metabolic responses to amino acid starvation in Candida albicans. *The EMBO journal* **21,** 5448-5456.

Venugopal, A., Bryk, R., Shi, S., Rhee, K., Rath, P., Schnappinger, D.*, et al.* (2011). Virulence of Mycobacterium tuberculosis depends on lipoamide dehydrogenase, a member of three multienzyme complexes. *Cell Host and Microbe* **9,** 21-31.

Wieland, H., Ullrich, S., Lang, F. and Neumeister, B. (2005). Intracellular multiplication of Legionella pneumophila depends on host cell amino acid transporter SLC1A5. *Molecular microbiology* **55,** 1528-1537.

Wood, H., Fehlner-Gardner, C., Berry, J., Fischer, E., Graham, B., Hackstadt, T.*, et al.* (2003). Regulation of tryptophan synthase gene expression in Chlamydia trachomatis. *Molecular microbiology* **49,** 1347-1359.

Yuk, J.-M., Shin, D.-M., Lee, H.-M., Yang, C.-S., Jin, H.S., Kim, K.-K.*, et al.* (2009). Vitamin D3 Induces Autophagy in Human Monocytes/Macrophages via Cathelicidin. *Cell Host and Microbe* **6,** 231-243.

Zelante, T., Fallarino, F., Bistoni, F., Puccetti, P. and Romani, L. (2009). Indoleamine 2,3-dioxygenase in infection: the paradox of an evasive strategy that benefits the host. *Microbes and infection / Institut Pasteur* **11,** 133-141.

Zhao, Z., Fux, B., Goodwin, M., Dunay, I.R., Strong, D., Miller, B.C.*, et al.* (2008). Autophagosome-independent essential function for the autophagy protein Atg5 in cellular immunity to intracellular pathogens. *Cell host & microbe* **4,** 458-469.

Section 1.4: Therapeutic Strategies for improving immune clearance of *Mycobacterium tuberculosis*

Overview

Despite the fact that treatment options for tuberculosis have existed for the last six decades, the disease remains one of the main global health threats. The need for new drugs is urgent, since shortening the current drug regimen would be beneficial, and since drug resistance cases are increasing in prevalence. After completely drying for many decades, the tuberculosis drug development pipeline has started to flow again, with one new FDA approved drug in the last year and many more in clinical trials. All of these potential therapeutics, however, target processes required for Mtb growth *in vitro,* likely due to the screening strategies employed to discover these growth-inhibiting compounds. Bacterial survival during infection, however, requires many other determinants, including targeted suppression of adaptive and innate immune responses, exploited host proteins used to create a more conducive milieu for the pathogen, and pathogen processes required to survive the stresses imposed by the host. Small molecules that induce protective host defenses or inhibit mycobacterial survival determinants will help the immune system kill Mtb more effectively. In order to explore the entire space of potential therapeutic strategies, we need to move beyond targeting essential genes and include virulence factors and host defenses as areas of chemical intervention. In this section, we review the mycobacterial adaptation to the host and discuss work that has led to the discovery of potential drugs that help improve the efficacy of host defenses,

either by inducing better (or more) host defense mechanisms or by making the bacterium more susceptible to existing defenses.

INTRODUCTION

Tuberculosis (TB), the disease caused by *Mycobacterium tuberculosis* (Mtb), has been treatable by antibiotics for over 50 years. Despite this, TB continues to be one of the deadliest infectious diseases in the world. In 2010, there were an estimated 8.8 million new cases of TB and 1.45 million deaths from TB worldwide[1]. Efforts aimed at global control of TB face multiple hurdles, including the long duration of treatment and the increasing incidence of drug resistant cases. Multidrug-resistant (MDR) TB caused an estimated 440,000 cases in 2010, and only 1% of these cases were treated with the regimen recommended by the World Health Organization[1]. New drugs are needed in the fight against TB to stem the tide of difficult-to-treat MDR cases and hopefully shorten the time needed to treat. Luckily, after a 45-year TB drug-development drought, a new TB drug, Bedaquiline, has been approved by the United States Food and Drug Administration, and multiple others are currently in Phase II or Phase III clinical trials[2, 3]. However, the long-term efficacy of these drugs depends on unknown factors such as their long-term toxicity and the speed with which new resistance develops. So it is crucial that the TB drug development pipeline does not dry up and that new drugs and new strategies for control continue to be identified.

When thinking about strategies for TB control, it is interesting to note that our immune systems usually do an adequate enough job to control disease. The adaptive immune system and the pathogen reaches an equilibrium, known clinically as latency, wherein disease is prevented but bacteria are not completely cleared[4]. Drugs that improve immunity or block pathogen strategies of survival

might be able to break this détente and spur pathogen clearance. Certainly in the 5-10% of infected individuals where immunity fails to prevent disease, potential immune-boosting strategies could be useful.

Most drugs target proteins that are essential for bacterial viability in laboratory culture conditions. Here, we highlight another strategy—therapeutics that help the immune system kill Mtb more effectively—and review promising targets and recent advances in drug development. There are two main ways to aid antimycobacterial immunity. First, we can design therapeutics to increase or induce effective immune control mechanisms. Interestingly, many of these strategies ultimately kill Mtb by putting the bacterium into lysosomes, overriding the key Mtb survival strategy of avoiding lysosomal delivery. Second, we can target Mtb proteins required for bacterial growth during infection—virulence factors and other proteins essential for its adaptation to the host—to make Mtb more susceptible to host immunity. By inhibiting mycobacterial processes that neutralize host defenses, we can turn these ineffective host mechanisms into strategies for control.

OVERVIEW OF HOST DEFENSES AND MTB RESPONSES

Mycobacterial adaptations to intracellular host defenses

Mycobacterium tuberculosis has adapted to survive its infectious niche and the pressures of host immunity. In this section, we provide a brief overview to immune-mediated stresses and Mtb adaptations to these stresses.

Mtb adaptations to the intracellular niche and innate immunity

Host vesicular trafficking is responsible for sequestering infecting microbes into lysosomes, where the pathogens are subjected to an arsenal of host defenses, including acid, oxidative stress, and antimicrobial peptides. Mtb is able to block this process, and during infection of macrophages in cell culture, Mtb survives primarily in unacidified phagosomes[5, 6]. Interestingly, the induction of autophagy can overcome this block and kill intracellular Mtb[7-10]. The exact mechanism by which Mtb blocks vesicular trafficking is unclear, but the many components required for this block have been elucidated. These include both bacterial virulence factors and exploited host proteins, and any of these determinants of phagolysosome fusion blockade could serve as potential therapeutic targets[6, 11, 12].

Mtb adaptations to metabolite starvation—scavenging and biosynthesis

While Mtb avoids many noxious agents by blocking delivery to lysosomes, it still faces metabolite starvation during infection. The intracellular environment seems to be relatively lacking in carbohydrate sources of carbon, and Mtb survives in the host by utilizing fatty acids. This is evidenced by the need for the glyoxylate

shunt, cholesterol catabolism and gluconeogenesis during infection of both macrophages and mice[13-17]. Furthermore, intracellular amino acid availability is limited during bacterial infection, and amino acid biosynthesis is active and required during Mtb infection[18-21]. These metabolic demands imposed by the immune system are successfully met by Mtb synthetic pathways, but blocking these pathways can restore the efficacy of immune-mediated starvation.

Mtb survival of reactive oxygen and nitrogen species

Reactive oxygen intermediates (ROI) and to a larger extent reactive nitrogen intermediates play important roles in limiting Mtb growth and maintaining latency in mice, but are also insufficient to clear bacteria[22-25]. Mtb can sense oxidative stress through a two component system called DosRS, and coordinate a protective response that is crucial for its survival during infection[26]. It employs an army of detoxifying enzymes and agents, protects from oxidatively damaged proteins using its proteasome, and repairs DNA, all processes that are required for survival during infection[26-29].

Mycobacterial adaptations to intracellular host defenses

Mtb manipulates the adaptive immune response through a myriad of different strategies. Through an unknown mechanism, it delays both antigen presentation to T cells in the lymph node and arrival of T cells to the locus of infection, thus slowing the initiation of adaptive immunity[30, 31]. Mtb also seems to manipulate adaptive immunity by limiting the expression of protective antigens, and overexpression of one particular antigen, Antigen 85B, improves CD4 activation and control of Mtb[32]. Furthermore, a study of Mtb genetic diversity revealed that

T cell antigens are more conserved than essential genes, suggesting that Mtb may have evolved ways to exploit T cells for its own survival[33]. Clever vaccine strategies that elicit protective responses at the expense of these Mtb-driven non-protective responses could be fruitful in shifting the host-pathogen equilibrium to the host's advantage. Mtb vaccine design is a rich and well-reviewed topic. While therapeutic strategies to boost adaptive immunity against Mtb are important, this review will focus on strategies that improve intracellular host defenses or increase Mtb susceptibility towards these defenses.

THERAPEUTIC STRATEGIES TO IMPROVE INTRACELLULAR HOST DEFENSES

Autophagy induction reduces mycobacterial survival in macrophages

Autophagy is an essential tool of cellular physiology. It occurs at a basal level in eukaryotic cells, but is especially important for maintaining homeostasis during certain physiologic stresses, such as starvation, organelle damage, or pathogen infection[34]. It was first identified as having anti-mycobacterial potential in a study that induced autophagy and observed decreased growth of *Mycobacterium bovis BCG* [7]. Later groups noticed that the anti-mycobacterial activity of two classes of IFN-γ-responsive GTPase, LRG-47 and Gbps, as well as the Vitamin D-induced antimicrobial peptide Cathelicidin, were all autophagy-dependent[9, 10, 35-37]. The formation of autophagosomes around Mtb and other bacterial pathogens, also called xenophagy, leads to lysosomal fusion and delivers Mtb to the toxic environment that it otherwise avoids[9, 19]. Thus, therapeutics that enhance autophagy could promote an effective host defense for Mtb control.

Autophagy-inducing peptides

Many existing therapeutics have autophagy-inducing effects, but these drugs all have autophagy-independent effects that may limit their clinical use as autophagy inducers[38]. Recently a group has developed a specific autophagy-inducing peptide that limits *Listeria monocytogenes* growth in cell culture models[39]. The peptide was derived from the Nef-binding region of an autophagy-related protein, Beclin1. In addition to its efficacy in inhibiting *L. monocytogenes* growth in macrophages, it also decreased replication of pathogenic viruses *in*

vivo[39]. The main advantage of an autophagy-specific therapeutic is the lack of off-target effects as would be seen with other autophagy-inducing drugs. The peptide was well tolerated by mice during the two-week-long daily treatment, with no histological damage seen in their vital organs[39]. This peptide is a promising candidate for TB treatment, and should be tested in a mouse model of Mtb infection.

Vitamin D

The role of autophagy in a host of different diseases has heightened the interest in developing autophagy-inducing drugs[38]. For inducing autophagy during TB, however, an inexpensive and safe treatment option might already exist. Vitamin D works through the antimicrobial peptide Cathelicidin to induce autophagy in Mtb-infected macrophages[10, 37, 40]. Along with IFN-γ, Vitamin D is able to kill both HIV and Mtb in macrophages in an autophagy-dependent manner[40]. Interestingly, the amount of Vitamin D in human serum varies substantially, and in one study, donated human serum with 48 nM of 25-hydroxyvitamin D3 was permissive of mycobacterial growth, while another human serum sample with 96 nM was able to kill 90% of mycobacteria in macrophages[37].

There is strong epidemiological evidence that low Vitamin D levels increase TB risk in human patients as well[41]. We get most of our Vitamin D through a synthetic process that begins in keratinocytes exposed to ultraviolet B radiation, although dietary sources of Vitamin D also exist. Thus, a lack of sunlight exposure and low dietary intake can easily result in Vitamin D insufficiency. Many cross-sectional and case-control studies, as well as a recent meta-analysis, have

linked low levels of Vitamin D with clinical TB[42-44]. Interestingly, groups in Cape Town and New York have also reported seasonal variability in TB incidence, as we might expect given that differences in sunlight exposure is one of the greatest sources of Vitamin D variability in the human population[43, 45]. In addition, polymorphisms that decrease activity of the Vitamin D Receptor (VDR), which governs immune cell responses to Vitamin D, also increases TB risk[44, 46].

The mounting evidence suggests a clear association between Vitamin D and TB risk, but a causal link, much less a directional one, has not yet been established. Cell culture models support the hypothesis that Vitamin D boosts immunity to TB, but reasonable alternative hypotheses exist, such as the possibility that TB causes ill patients to stay indoors and lose sunlight exposure. Randomized control trials (RCTs) assessing Vitamin D as adjunctive therapy have mostly failed to demonstrate an effect in the primary treatment outcomes[41, 47-50]. Each study looked to measure either faster or increased sputum culture clearance in their measurement arms, but none found a statistically significant improvement. Coussens et al. found a small effect in a subgroup with a recessive allele of VDR, but it is important to note that randomization is compromised in subgroup analyses, making them difficult to interpret[47].

Vitamin D has a clear anti-mycobacterial effect in cell culture models, and observational epidemiology has proven a clear link between Vitamin D deficiency and TB risk. The failure of RCTs so far to show clinical efficacy of Vitamin D during treatment is unfortunate, but hardly spells the end of Vitamin D as a potential tool against TB in the future. Firstly, the RCTs were attempting to

measure improvement in clinical outcome beyond that seen in first-line treatment regimens. The effects were likely to be small, and these studies were not powered to find small effects. Secondly, the cell biology models of suggest that Vitamin D works best against Mtb contained in macrophages. While this is true of many bacilli in active disease, Mtb containment within macrophages is most evident early in infection, and it is possible that we should be investigating Vitamin D as a nutritional supplement for prevention rather than for treatment. To perform such a study, even amongst HIV positive patients, would require huge cohorts. But while the costs would be huge, the upside of proving that a cheap dietary supplement could prevent TB would be tremendous.

Inhibition of host proteins exploited by Mtb

Mtb exploits multiple host proteins for survival[11, 51], and inhibiting these proteins could remove an important survival tool for Mtb during infection.

Akt kinase inhibitor

Intracellular pathogens require many host factors in order to replicate in the host cell. Realizing this, researchers performed both genetic and chemical screens to identify these exploited host factors. One screen done in *Salmonella* yielded a host target and chemical inhibitor that also worked in for Mtb infection[52]. Using a kinase inhibitor library, the authors found that one compound, HA-89 effectively inhibits both *Salmonella* and Mtb growth in cell culture. They tested HA-89 against an in vitro panel of human kinases and determined that the only kinase inhibited by HA-89 that as required for *Salmonella* growth in their genetic screen was Akt. Of note, Akt was also found to be required for Mtb growth by an

independent screen[11, 51]. While this is suggestive of Akt being the target, inhibition of a bacterial kinase (Mtb expresses two essential serine/threonine kinases) or another human kinase could also be responsible for the effect on bacterial growth. The mechanism by which Akt supports mycobacterial growth is unclear, but the authors suggest a link to vesicular trafficking and phagolysosome fusion. Akt inhibitors have been used in mice to treat cancers, but at least a few reports have noted general toxicity. Regardless, inhibition of Akt is a proof of concept for targeting exploited host proteins, and efficacy studies in mice would provide additional insight into Akt as a potential target in TB therapy.

Imatinib

Imatinib was one of the first rationally-designed drugs for cancer, and was made to inhibit the Abl kinase, which is activated by a chromosomal translocation in chromic myeloid leukemia[53]. Investigators have recently shown that imatinib and another Abl inhibitor decrease Mtb growth in macrophages and in mice[54]. Imatinib treatment of inhibited macrophages also increases bacterial colocalization with lysosomal markers, suggesting that Abl is exploited by Mtb to affect vesicular trafficking[54, 55]. Furthermore, Mtb killing with imatinib is dependent on the vacuolar ATPase and acidification of the phagosome. Interestingly, imatinib also has an effect on mycobacterial growth in Abl knockout macrophages, suggesting that it has non-Abl targets as well.

None of the therapeutics mentioned so far have been tested in model systems in the presence of first-line TB drugs, and so interactions with existing treatments is

an unknown, but crucial piece of the puzzle. There is no *a priori* assumption; many of the potential therapeutics discussed so far change the intracellular milieu of Mtb, and this change in environment could improve or hamper the activity of current anti-mycobacterials. In this study, interaction with an existing drug, rifampin, is directly addressed. Interestingly, the authors show that imatinib acts synergistically with rifampin, though the reason is unclear. This is a clear argument for modeling the interaction of new therapies with existing ones. This is especially true with the molecules in this review; since they affect the host-pathogen interface, many result in driving the bacterium out of its self-crafted and relatively bacteria-friendly niche, and into a more stressful environment where existing drugs could have greater efficacy.

Summary

Mtb lives in an infectious niche that it is tolerates perfectly, and for the most part, our bodies tolerate Mtb without any consequences to our health. However, this equilibrium permits the survival of bacilli that could cause disease in the future. A small nudge in this equilibrium could provide the tipping point needed to improve disease and treatment outcomes. In addition to the ideas and potential therapeutics outlined above, there are many other possibilities. Arginine, a compound that theoretically increases nitrc oxide, has been tested in infected people, although the results from those studies have been largely inconclusive[56]. Important cytokines have also been used to treat TB in animal models and in humans. Recombinant cytokines, however, remain fairly expensive and thus difficult to imagine as a strategy against global TB. Furthermore, many of the

host-friendly cytokines, such as IFN-γ, are already at a high enough level upon infection, and there are possibly decreasing margins-of-return with what amounts to an incremental increase. But while recombinant proteins or other biologics might not be feasible, small molecule agonists of pattern recognition receptors have been tailored to increase cytokine release, and these might prove useful in the future.

THERAPEUTIC STRATEGIES THAT MAKE MTB MORE SUSCEPTIBLE TO HOST DEFENSES

Bacterial pathogens and their hosts have engaged in an evolutionary arms race, and it's likely that both have weapons in their arsenal that have been rendered ineffective by the opposing side. If we can understand and block the mechanisms that TB uses to neutralize host defenses, we can turn an otherwise ineffective host process into a pathway that kills Mtb. In this section, we will review three broad categories of mycobacterial adaptations to the host environment. First, we will discuss inhibitors of the determinants of the phagolysosome block. Second, we will review inhibitors of mycobacterial protection against nitric oxide. Finally, we will review small molecules that take aim at infection-specific metabolism.

Targeting mycobacterial mechanisms that inhibit phagolysosome fusion

PtpA and PtpB

Of the many mycobacterial factors that are purported to be involved blocking phagolysosome fusion, the secreted phosphatases have attracted the most attention as targets for chemical inhibition[12]. Mtb expresses two secreted phosphatases, PtpA and PtpB[57]. Both are required for survival in macrophages[58, 59], although only PtpB is required for growth in an animal model[60, 61]. Both phosphatases were first identified through genome sequencing and their similarity to eukaryotic low molecular weight phosphatases (LMW-PTP). Furthermore, despite the fact that neither phosphatase has a recognizable secretion signal, both phosphatases can be found in culture filtrates, and antibodies against PtpA can detect its presence in macrophages during a cell

culture infection[57, 58, 62]. Upon secretion into the host cell, PtpA dephosphorylates VPS33B and excludes the vATPase proton pump from the phagosome membrane, preventing phagosome acidification[58, 63]. Host-pathogen protein interactions are difficult to study in physiological conditions, as the pathogen protein often has to be overexpressed in the host cell in order to perform immunoprecipitation-Western blots or fluorescent microscopy with sufficient resolution. In the case of PtpA, however, genetic evidence also supports the importance of its interaction with the vATPase. A phosphatase-competent point mutant that cannot interact with vATPase has impaired growth in macrophages and ends up in acidic phagosomes[63]. The target of PtpB has not yet been established, but expression of PtpB in macrophages suggests that it may activate kinase pathways that eventually lead to decreased IL-6 secretion[59]. It has also been shown in vitro to have phosphatase activity at both serine/threonine and tyrosine residues[64]. But while intracellular secretion and activity of PtpB is not established, it has been shown to be required for growth in an animal model[60, 61].

There are many attractive things about targeting these phosphatases for therapeutic intervention. Both are required for macrophage growth; while similar to eukaryotic phosphatases, neither has a human ortholog, suggesting that target-specificity is achievable with small molecule inhibitors; and secretion into the host might make them more exposed to inhibitors, which do not have to cross the formidable barrier of the mycobacterial cell wall. There have been many reports of PtpA inhibitors, though the best molecules have been from a series of chalcones with K_i values ranging from 4.9 to 21.3 μM *in vitro*[65-67]. Importantly, the

chalcones have been tested in macrophages and have been shown to be potent inhibitors of Mtb growth in cells[67]. Chemical inhibition of PtpB activity in vitro has been studied by many groups, but only a few have tested its efficacy on live bacteria[59, 64, 65, 68-75]. The structure of PtpB has been used as used as a scaffold for informed drug design, and some groups have even used the existence of a secondary site to design bidentate inhibitors, resulting in some of the most potent inhibitors to date[69, 70]. Efficacy studies in macrophages have been done for only a small fraction of the compounds with known *in vitro* activity, but a few promising molecules, including two dual-site inhibitors, an isoxazole derivative and a salicylate-based compound, can significantly decrease mycobacterial growth in macrophages[59, 64].

The large number of *in vitro* inhibitors of these enzymes is encouraging, but many challenges lie ahead. Since PtpA is non-essential in mice, there is no good animal model for further testing of PtpA inhibitors. However, PtpB is essential in guinea pigs, and inhibitor efficacy in this model would be critical for determining the true promise of this phosphatase as a drug target.

PknG

Protein Kinase G (PknG) is a kinase that appears to help support the growth of *Mycobacterium bovis BCG* in macrophages by blocking lysosomal fusion[76-79]. Interestingly, Mtb PknG appears to be either essential or at least required for optimal growth *in vitro*, so its role in Mtb compared to its role in BCG is unclear[80]. Small molecule inhibitors, the best of which is a tetrahydrobenzothiophene called AX20017, of PknG that limit both Mtb and BCG growth in macrophages have

been described[76]. While these inhibitors have been tested against a panel of bacterial and human kinases, it is worth noting that the *in vivo* target has not been definitively characterized, and AX20017 has not been tested in mice[81].

Targeting Mtb proteins that guard against oxidative stress

Reactive oxygen and nitrogen species are an integral part of the host response against Mtb, especially in mice[22-26]. Damage mediated by ROI and RNI is complex, and Mtb encodes multiple sensing and damage response pathways to help it survive oxidative damage.

RNI detoxification

Sulfate transport and reduction as well as cysteine biosynthesis is upregulated and during dormancy and required in long-term survival models[27]. Interestingly, cysteine auxotrophs are attenuated in wildtype mice, but almost fully virulent in *Rag*$^{-/-}$ mice, which lack adaptive immunity[82]. The mechanism for this adaptive immunity-driven requirement for cysteine is not fully characterized, but it is thought to relate to the need for cysteine in the synthesis of mycothiol, a reducing agent that protects the bacterium from oxidative damage. Inhibitors of cysteine biosynthesis have been identified, and it would be interesting to test these inhibitors in infection models[83, 84].

Proteasome inhibitors

A screen for mycobacterial survival determinants in nitric oxide (NO) revealed that the mycobacterial proteasome was required for NO resistance and survival during mouse infection[28]. Proteasome knock strains were later shown to be required for persistence in the later stages of mouse infection as well[85].

Interestingly, when these same strains were used to infect mice lacking the inducible nitric oxide synthase (iNOS), the expected rescue of virulence was not seen, suggesting that the proteasome might be important for additional reasons *in vivo*[28]. Nevertheless, the mycobacterial proteasome remained a promising target for drug development since proteasomes from other species had been successfully targeted. In 2009 an inhibitor was described that was selective for the mycobacterial proteasome and had very little activity against the human proteasome, demonstrating that species-specificity was possible[86]. The same group also described a more potent proteasome inhibitor that did have cross-species activity, showing that selectivity is a major issue when designing proteasome drugs[87].

DNA damage response inhibitors

Another hit from the NO survival screen was DNA damage repair. UvrB, a component of the nucleotide excision repair machinery, was also found to be important for surviving NO-stress[28]. UvrB knockouts are attenuated in mice and macrophages[29]. Growth of the UvrB mutants is somewhat restored in $iNOS^{-/-}$ mice, although increased growth is no larger an effect than seen in wt Mtb[29]. Recently, a whole-cell screen for small molecules inhibitors that resulted in hypersusceptibility to UV resulted in the discovery of UvrB inhibitors[88]. While the effects of the inhibitor on NO survival was not shown, this study demonstrates that DNA excision repair can be targeted. Coupled with genetic evidence for its need during infection, UvrB inhibition could be a promising target moving forward.

Targeting mycobacterial metabolism during infection

It is thought that one of the advantages of being an intracellular pathogen is the abundance of nutrients inside host cells, since the host itself needs a steady supply of energy-rich metabolites[89]. However, it is clear that the metabolic demands of Mtb *in vitro* and *in vivo* are very different, and that immune-mediated starvation can drive this change in metabolic requirements[90]. Targeting these metabolic pathways then, can help turn the immune system's futile starvation attempts into real mechanisms of Mtb control.

Glyoxylate shunt inhibitors

Isocitrate lyase (Icl) and the glyoxylate shunt are required for mycobacterial growth in macrophages and mice, presumably because fatty acids are needed as a carbon source because glycolytic sources are unavailable or glycolysis intermediates are being shunted away from the TCA cycle[13, 14]. One of the first reports of Icl essentiality in vivo also presented a chemical inhibitor, which decreased growth in macrophages, albeit at high, millimolar concentrations[14]. The authors note that the active site of Icl is very polar and small, and is thus a relatively difficult enzyme to inhibit. There are other reports of Icl enzymatic inhibition, but interestingly many of these have low MIC values *in vitro*. Since Icl is not essential for Mtb growth in vitro, these molecules likely have off-target effects that limit mycobacterial growth[91-93]. While Icl seems to be a poor target for chemical inhibition, the second enzyme in the pathway, malate synthase (GlcB), has an active site structure more conducive to drug design. While there are conflicting reports on whether or not GlcB is required for survival *in vitro*, its inhibition should terminate flux through the *in vivo* essential glyoxylate shunt.

Recently, a GlcB inhibitor has been described that had a 2-log reduction in mycobacterial growth after 9 days of infection and treatment, showing early promise as a possible therapeutic strategy[16].

Amino Acid biosynthesis inhibitors

Amino acid biosynthesis is required for Mtb growth *in vivo*, and it has been shown in many other bacterial infection models that host cells are able to starve intracellular bacteria of amino acids[18-21]. Amino acid biosynthesis inhibitors have been utilized in many other organisms, and researchers have speculated that the shikimate pathway (responsible for the synthesis of aromatic amino acids), which has been successfully targeted in other organisms, might be a good target for Mtb[94, 95].

CONCLUDING REMARKS

The growing problem of drug resistance demands new drugs and new targets for the global fight against TB. One of the key challenges in developing truly new drugs is identifying new pathways to target. Expanding the list beyond proteins required for growth in laboratory conditions can help. In this review, we have given an overview of current and future strategies aimed at helping the immune system kill Mtb more effectively. The therapeutics described include compounds that target host proteins, inducers of host defenses, and inhibitors of bacterial survival determinants during infection. Also key in TB drug design are the overlapping challenges of decreasing treatment time and killing non-replicating bacteria. Although non-replicating populations are complex and poorly understood, it is not difficult to imagine that drugs designed to limit bacterial growth *in vitro* might not be optimized for killing non-replicating bacteria. While this will vary case by case, it is possible that treatments aiming to boost immune clearance of bacteria might have better efficacy against these difficult bacterial subpopulations, and thus accelerate treatment.

Section 1.5: Bibliography (II)

1. WHO. (http://www.who.int/tb/publications/global_report/2011/en/, 2011).

2. Osborne, R. First novel anti-tuberculosis drug in 40 years. *Nat Biotechnol* **31**, 89-91 (2013).

3. Zumla, A., Hafner, R., Lienhardt, C., Hoelscher, M. & Nunn, A. Advancing the development of tuberculosis therapy. *Nat Rev Drug Discov* **11**, 171-2 (2012).

4. Ernst, J.D. The immunological life cycle of tuberculosis. *Nat Rev Immunol* **12**, 581-91 (2012).

5. Clemens, D.L. & Horowitz, M.A. Characterization of the Mycobacteriummberctdosis Phagosome and Evidence that Phagosomal Maturation Is Inhibited. *Journal of Expermental Medicine* **181**, 257-270 (1995).

6. Pethe, K. et al. Isolation of Mycobacterium tuberculosis mutants defective in the arrest of phagosome maturation. *Proc Natl Acad Sci U S A* **101**, 13642-7 (2004).

7. Gutierrez, M.G. et al. Autophagy is a defense mechanism inhibiting BCG and Mycobacterium tuberculosis survival in infected macrophages. *Cell* **119**, 753-66 (2004).

8. Alonso, S., Pethe, K., Russell, D.G. & Purdy, G.E. Lysosomal killing of Mycobacterium mediated by ubiquitin-derived peptides is enhanced by autophagy. *Proc Natl Acad Sci U S A* **104**, 6031-6 (2007).

9. Singh, S.B., Davis, A.S., Taylor, G.A. & Deretic, V. Human IRGM induces autophagy to eliminate intracellular mycobacteria. *Science* **313**, 1438-41 (2006).

10. Yuk, J.M. et al. Vitamin D3 induces autophagy in human monocytes/macrophages via cathelicidin. *Cell Host Microbe* **6**, 231-43 (2009).

11. Kumar, D. et al. Genome-wide analysis of the host intracellular network that regulates survival of Mycobacterium tuberculosis. *Cell* **140**, 731-43 (2010).

12. Wong, D., Chao, J.D. & Av-Gay, Y. Mycobacterium tuberculosis-secreted phosphatases: from pathogenesis to targets for TB drug development. *Trends Microbiol* **21**, 100-9 (2013).

13. McKinney, J.D. et al. Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**, 735-738 (2000).

14. Munoz-Elias, E.J. & McKinney, J.D. Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nat Med* **11**, 638-44 (2005).

15. Ouellet, H., Johnston, J.B. & de Montellano, P.R. Cholesterol catabolism as a therapeutic target in Mycobacterium tuberculosis. *Trends Microbiol* **19**, 530-9 (2011).

16. Krieger, I.V. et al. Structure-guided discovery of phenyl-diketo acids as potent inhibitors of M. tuberculosis malate synthase. *Chem Biol* **19**, 1556-67 (2012).

17. Pandey, A.K. & Sassetti, C.M. Mycobacterial persistence requires the utilization of host cholesterol. *Proc Natl Acad Sci U S A* **105**, 4376-80 (2008).

18. Smith, D.A., Parish, T., Stoker, N.G. & Bancroft, G.J. Characterization of auxotrophic mutants of Mycobacterium tuberculosis and their potential as vaccine candidates. *Infection and immunity* **69**, 1142-1150 (2001).

19. Pavelka, M.S., Chen, B., Kelley, C.L., Collins, F.M. & Jacobs Jr, W.R. Vaccine efficacy of a lysine auxotroph of Mycobacterium tuberculosis. *Infection and immunity* **71**, 4190-4192 (2003).

20. Hondalus, M.K. et al. Attenuation of and protection induced by a leucine auxotroph of Mycobacterium tuberculosis. *Infection and immunity* **68**, 2888-2898 (2000).

21. Tattoli, I. et al. Amino Acid Starvation Induced by Invasive Bacterial Pathogens Triggers an Innate Host Defense Program. *Cell Host and Microbe* **11**, 563-575 (2012).

22. Flynn, J.L., Scanga, C.A., Tanaka, K.E. & Chan, J. Effects of Aminoguanidine on Latent Murine Tuberculosis. *Journal of immunology (Baltimore, Md : 1950)* **160**, 1796-1803 (1998).

23. Macmicking, J.D. et al. Identification of nitric oxide synthase as a protective locus against tuberculosis. *Proceedings of the National*

Academy of Sciences of the United States of America **94**, 5243–5248 (1997).

24. Cooper, A.M., Segal, B.H., Frank, A.A., Holland, S.M. & Orme, I.M. Transient Loss of Resistance to Pulmonary Tuberculosis in p47 phox-/- Mice. *Infection and immunity* **68**, 1231-1234 (2000).

25. Jung, Y.J., LaCourse, R., Ryan, L. & North, R.J. Virulent but not Avirulent Mycobacterium tuberculosis Can Evade the Growth Inhibitory Action of a T Helper 1-dependent, Nitric Oxide Synthase 2-independent Defense in Mice. *Journal of Experimental Medicine* **196**, 991-998 (2002).

26. Trivedi, A., Singh, N., Bhat, S.A., Gupta, P. & Kumar, A. Redox biology of tuberculosis pathogenesis. *Adv Microb Physiol* **60**, 263-324 (2012).

27. Ehrt, S. & Schnappinger, D. Mycobacterial survival strategies in the phagosome: defence against host stresses. *Cell Microbiol* **11**, 1170-8 (2009).

28. Darwin, K.H., Ehrt, S., Gutierrez-Ramos, J.C., Weich, N. & Nathan, C.F. The proteasome of Mycobacterium tuberculosis is required for resistance to nitric oxide. *Science* **302**, 1963-6 (2003).

29. Darwin, K.H. & Nathan, C.F. Role for nucleotide excision repair in virulence of Mycobacterium tuberculosis. *Infect Immun* **73**, 4581-7 (2005).

30. Wolf, A.J. et al. Initiation of the adaptive immune response to Mycobacterium tuberculosis depends on antigen production in the local lymph node, not the lungs. *J Exp Med* **205**, 105-15 (2008).

31. Shafiani, S., Tucker-Heard, G., Kariyone, A., Takatsu, K. & Urdahl, K.B. Pathogen-specific regulatory T cells delay the arrival of effector T cells in the lung during early tuberculosis. *J Exp Med* **207**, 1409-20 (2010).

32. Bold, T.A., Banaei, N., Wolf, A.J. & Ernst, J.D. Suboptimal Activation of Antigen-Specific CD4 Effector Cells Enables Presistence of *M. tuberculosis* In Vivo. *PLoS Pathogens* **7**, e1002063 (2011).

33. Comas, I. et al. Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. *Nat Genet* **42**, 498-503 (2010).

34. Levine, B. Eating oneself and uninvited guests: autophagy-related pathways in cellular defense. *Cell* **120**, 159-62 (2005).

35. Kim, B.H. et al. A family of IFN-gamma-inducible 65-kD GTPases protects against bacterial infection. *Science* **332**, 717-21 (2011).

36. Singh, S.B. et al. Human IRGM regulates autophagy and cell-autonomous immunity functions through mitochondria. *Nat Cell Biol* **12**, 1154-65 (2010).

37. Fabri, M. et al. Vitamin D is required for IFN-gamma-mediated antimicrobial activity of human macrophages. *Sci Transl Med* **3**, 104ra102 (2011).

38. Rubinsztein, D.C., Codogno, P. & Levine, B. Autophagy modulation as a potential therapeutic target for diverse diseases. *Nat Rev Drug Discov* **11**, 709-30 (2012).

39. Shoji-Kawata, S. et al. Identification of a candidate therapeutic autophagy-inducing peptide. *Nature* **494**, 201-6 (2013).

40. Campbell, G.R. & Spector, S.A. Vitamin D inhibits human immunodeficiency virus type 1 and Mycobacterium tuberculosis infection in macrophages through the induction of autophagy. *PLoS Pathog* **8**, e1002689 (2012).

41. Ralph, A.P., Lucas, R.M. & Norval, M. Vitamin D and solar ultraviolet radiation in the risk and treatment of tuberculosis. *The Lancet Infectious Diseases* **13**, 77-88 (2013).

42. Nnoaham, K.E. & Clarke, A. Low serum vitamin D levels and tuberculosis: a systematic review and meta-analysis. *Int J Epidemiol* **37**, 113-9 (2008).

43. Martineau, A.R. et al. Reciprocal seasonal variation in vitamin D status and tuberculosis notifications in Cape Town, South Africa. *Proc Natl Acad Sci U S A* **108**, 19013-7 (2011).

44. Wilkinson, R.J. et al. Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London: a case-control study. *The Lancet* **355**, 618-621 (2000).

45. Parrinello, C.M., Crossa, A. & Harris, T.G. Seasonality of tuberculosis in New York City, 1990-2007. *Int J Tuberc Lung Dis* **16**, 32-7 (2012).

46. Gao, L., Tao, Y., Zhang, Q. & Jin, Q. Vitamin D receptor genetic polymorphisms and tuberculosis: updated systematic review and meta-analysis. *Int J Tuberc Lung Dis* **14**, 15-23 (2010).

47. Coussens, A.K. et al. Vitamin D accelerates resolution of inflammatory responses during tuberculosis treatment. *Proc Natl Acad Sci U S A* **109**, 15449-54 (2012).

48. Salahuddin, N. et al. Vitamin D accelerates clinical recovery from tuberculosis: results of the SUCCINCT Study [Supplementary Cholecalciferol in recovery from tuberculosis]. A randomized, placebo-controlled, clinical trial of vitamin D supplementation in patients with pulmonary tuberculosis'. *BMC Infect Dis* **13**, 22 (2013).

49. Wejse, C. et al. Vitamin D as supplementary treatment for tuberculosis: a double-blind, randomized, placebo-controlled trial. *Am J Respir Crit Care Med* **179**, 843-50 (2009).

50. Martineau, A.R. et al. A single dose of vitamin D enhances immunity to mycobacteria. *Am J Respir Crit Care Med* **176**, 208-13 (2007).

51. Jayaswal, S. et al. Identification of host-dependent survival factors for intracellular Mycobacterium tuberculosis through an siRNA screen. *PLoS Pathog* **6**, e1000839 (2010).

52. Kuijl, C. et al. Intracellular bacterial growth is controlled by a kinase network around PKB/AKT1. *Nature* **450**, 725-30 (2007).

53. Deninger, M., Buchdunger, E. & Druker, B.J. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood* **105**, 2640-2653 (2005).

54. Napier, R.J. et al. Imatinib-sensitive tyrosine kinases regulate mycobacterial pathogenesis and represent therapeutic targets against tuberculosis. *Cell Host Microbe* **10**, 475-85 (2011).

55. Bruns, H. et al. Abelson tyrosine kinase controls phagosomal acidification required for killing of Mycobacterium tuberculosis in human macrophages. *J Immunol* **189**, 4069-78 (2012).

56. Schon, T. et al. Arginine as an adjuvant to chemotherapy improves clinical outcome in active tuberculosis. *European Respiratory Journal* **21**, 483-488 (2003).

57. Koul, A. et al. Cloning and Charatcerization of Secretory Tyrosine Phosphatases of *Mycobacterium tubercuosis*. *Journal of Bacteriology* **182**, 5425 (2000).

58. Bach, H., Papavinasasundaram, K.G., Wong, D., Hmama, Z. & Av-Gay, Y. Mycobacterium tuberculosis virulence is mediated by PtpA dephosphorylation of human vacuolar protein sorting 33B. *Cell Host Microbe* **3**, 316-22 (2008).

59. Zhou, B. et al. Targeting mycobacterium protein tyrosine phosphatase B for antituberculosis agents. *Proc Natl Acad Sci U S A* **107**, 4573-8 (2010).

60. Singh, R. et al. Disruption of mptpB impairs the ability of Mycobacterium tuberculosis to survive in guinea pigs. *Molecular microbiology* **50**, 751-762 (2003).

61. Grundner, C., Cox, J.S. & Alber, T. Protein tyrosine phosphatase PtpA is not required for Mycobacterium tuberculosis growth in mice. *FEMS Microbiol Lett* **287**, 181-4 (2008).

62. Cowley, S.C., Babakaiff, R. & Av-Gay, Y. Expression and localization of *Mycobacterium tuberculosis* protein tyrosine phosphatase PtpA. *Research in Microbiology* **153**, 233-241 (2002).

63. Wong, D., Bach, H., Sun, J., Hmama, Z. & Av-Gay, Y. Mycobacterium tuberculosis protein tyrosine phosphatase (PtpA) excludes host vacuolar-H+-ATPase to inhibit phagosome acidification. *Proc Natl Acad Sci U S A* **108**, 19371-6 (2011).

64. Beresford, N.J. et al. Inhibition of MptpB phosphatase from Mycobacterium tuberculosis impairs mycobacterial survival in macrophages. *J Antimicrob Chemother* **63**, 928-36 (2009).

65. Chiaradia, L.D. et al. Synthesis, biological evaluation, and molecular modeling of chalcone derivatives as potent inhibitors of Mycobacterium tuberculosis protein tyrosine phosphatases (PtpA and PtpB). *J Med Chem* **55**, 390-402 (2012).

66. Chiaradia, L.D. et al. Synthetic chalcones as efficient inhibitors of Mycobacterium tuberculosis protein tyrosine phosphatase PtpA. *Bioorg Med Chem Lett* **18**, 6227-30 (2008).

67. Mascarello, A. et al. Inhibition of Mycobacterium tuberculosis tyrosine phosphatase PtpA by synthetic chalcones: kinetics, molecular modeling, toxicity and effect on growth. *Bioorg Med Chem* **18**, 3783-9 (2010).

68. Soellner, M., Rawls, K.A., Grundner, C., Alber, T. & Ellman, J.A. Fragment-Based Substrate Activity Screening Method for the Identification of Potent Inhibitors of the Mycobacterium tuberculosis Phosphatase PtpB. *Journal of the American Chemical Society* **129**, 9613-9615 (2007).

69. Grundner, C. et al. Structural basis for selective inhibition of Mycobacterium tuberculosis protein tyrosine phosphatase PtpB. *Structure* **15**, 499-509 (2007).

70. Tan, L.P. et al. Mycobacterium tuberculosis Protein Tyrosine Phosphatase B (MptpB) Inhibitors Using Click Chemistry. *Organic Letters* **11**, 5102-5101 (2009).

71. Rawls, K.A., Grundner, C. & Ellman, J.A. Design and synthesis of nonpeptidic, small molecule inhibitors for the Mycobacterium tuberculosis protein tyrosine phosphatase PtpB. *Org Biomol Chem* **8**, 4066-70 (2010).

72. Chen, L. et al. Identification and characterization of novel inhibitors of mPTPB, an essential virulent phosphatase from Mycobacterium tuberculosis. *ACS Med Chem Lett* **1**, 355-359 (2010).

73. He, R. et al. Double click reaction for the acquisition of a highly potent and selective mPTPB inhibitor. *ChemMedChem* **5**, 2051-6 (2010).

74. He, Y. et al. Discovery and Evaluation of Novel Inhibitors of Mycobacterium Protein Tyrosine Phosphatase B from the 6-Hydroxy-benzofuran-5-carboxylic Acid Scaffold. *J Med Chem* **56**, 832-42 (2013).

75. He, R. et al. Organocatalytic multicomponent reaction for the acquisition of a selective inhibitor of mPTPB, a virulence factor of tuberculosis. *Chem Commun (Camb)* **49**, 2064-6 (2013).

76. Walburger, A. et al. Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science* **304**, 1800-4 (2004).

77. Sassetti, C.M., Boyd, D.H. & Rubin, E.J. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular microbiology* **48**, 77-84 (2003).

78. Zhang, Y.J. et al. Global assessment of genomic regions required for growth in Mycobacterium tuberculosis. *PLoS Pathog* **8**, e1002946 (2012).

79. DeJesus, M. et al. Bayesian Analysis of Gene Essentiality based on Sequencing of Transposon Insertion Libraries. *Bioinformatics* **10**, 1093 (2013).

80. Cowley, S. et al. The Mycobacterium tuberculosis protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo. *Mol Microbiol* **52**, 1691-702 (2004).

81. Scherr, N. et al. Structural basis for the specific inhibition of protein kinase G, a virulence factor of Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12151-12156 (2007).

82. Senaratne, R.H. et al. 5'-Adenosinephosphosulphate reductase (CysH) protects Mycobacterium tuberculosis against free radicals during chronic infection phase in mice. *Mol Microbiol* **59**, 1744-53 (2006).

83. Schnell, R., Oehlmann, W., Singh, M. & Schneider, G. Structural insights into catalysis and inhibition of O-acetylserine sulfhydrylase from Mycobacterium tuberculosis. Crystal structures of the enzyme alpha-aminoacrylate intermediate and an enzyme-inhibitor complex. *J Biol Chem* **282**, 23473-81 (2007).

84. Jean Kumar, V.U. et al. Discovery of novel inhibitors targeting the Mycobacterium tuberculosis O-acetylserine sulfhydrylase (CysK1) using virtual high-throughput screening. *Bioorg Med Chem Lett* **23**, 1182-6 (2013).

85. Gandotra, S., Schnappinger, D., Monteleone, M., Hillen, W. & Ehrt, S. In vivo gene silencing identifies the Mycobacterium tuberculosis proteasome as essential for the bacteria to persist in mice. *Nat Med* **13**, 1515-20 (2007).

86. Lin, G. et al. Inhibitors selective for mycobacterial versus human proteasomes. *Nature* **461**, 621-6 (2009).

87. Lin, G., Li, D., Chidawanyika, T., Nathan, C. & Li, H. Fellutamide B is a potent inhibitor of the Mycobacterium tuberculosis proteasome. *Arch Biochem Biophys* **501**, 214-20 (2010).

88. Mazloum, N. et al. Identification of a chemical that inhibits the mycobacterial UvrABC complex in nucleotide excision repair. *Biochemistry* **50**, 1329-35 (2011).

89. Rohmer, L., Hocquet, D. & Miller, S.I. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends Microbiol* **19**, 341-8 (2011).

90. Rhee, K.Y. et al. Central carbon metabolism in Mycobacterium tuberculosis: an unexpected frontier. *Trends Microbiol* **19**, 307-14 (2011).

91. Shingnapurkar, D. et al. Synthesis and characterization of pyruvate-isoniazid analogs and their copper complexes as potential ICL inhibitors. *Bioorg Med Chem Lett* **22**, 3172-6 (2012).

92. Ji, L., Long, Q., Yang, D. & Xie, J. Identification of Mannich Base as a Novel Inhibitor of Mycobacterium Tuberculosis Isocitrate by High-Throughput Screening. *International Journal of Biological Sciences* **7**, 376-382 (2011).

93. Sriram, D. et al. Synthesis of various 3-nitropropionamides as Mycobacterium tuberculosis isocitrate lyase inhibitor. *Bioorg Med Chem Lett* **21**, 5149-54 (2011).

94. Dias, M.V. et al. Structural investigation of inhibitor designs targeting 3-dehydroquinate dehydratase from the shikimate pathway of Mycobacterium tuberculosis. *Biochem J* **436**, 729-39 (2011).

95. Reichau, S. et al. Potent inhibitors of a shikimate pathway enzyme from Mycobacterium tuberculosis: combining mechanism- and modeling-based design. *J Biol Chem* **286**, 16197-207 (2011).

Chapter 2.

Genomic Regions Required for *Mycobacterium tuberculosis* Growth

This chapter contains two published manuscripts:

Zhang YJ, Ioerger TR, Huttenhower C, Long JE, Sassetti CM, Sacchettini JC, Rubin, EJ. Global assessment of genomic regions required for growth in Mycobacterium tuberculosis. *PLoS Pathog* **8**, e1002946 (2012).

Dejesus M, Zhang YJ, Sassetti CM, Rubin EJ, Sacchettini JC, et al. Bayesian Analysis of Gene Essentiality based on Sequencing of Transposon Insertion Libraries. Bioinformatics. doi:10.1093/bioinformatics/btt043 (2013).

Section 2.1: Overview and Attributions

Overview. Discovering genes required for bacterial growth is foundational for both basic and applied biology. It provides key biological insight into bacterial physiology and identifies potential drug targets. This chapter contains two manuscripts that define genetic requirements for *Mycobacterium tuberculosis* (Mtb) growth in laboratory culture conditions. Both manuscripts analyze the same data—the number of insertions sequenced from an Mtb transposon mutant library grown on the standard solid rich-media in the field, 7H10 supplemented with glycerol. Previous analyses of transposon insertion counts from deep-sequencing focused on measuring the underrepresentation of insertion counts across annotated genes. Both manuscripts tackle an important oversight in this strategy: genes are often multi-component units, and the individual units could be differentially required. The first manuscript treats the genome as a series of random windows, and analyzes the requirement of each of these random windows. Most genes are composed completely of non-required windows; a smaller fraction is composed entirely of required windows. This paper highlights a third possibility—genes that contain both required and non-required windows. These genes could be representative of multimodular proteins wherein only some domains are required for growth, or non-required protein-coding genes that encode a required antisense RNA on the complement strand. Importantly, these are required genetic elements that are missed when genes are assessed for requirement as a single unit. The second paper also acknowledges that essential genes might be essential only because of a short essential stretch of coding

sequence. The task, then, is to find essential stretches of genetic code regardless of their length relative to the gene as a whole. By looking for continuous runs of potential insertion sites that lack insertions, this paper defines an essential gene as a gene that contains a statistically significant stretch of insertion-less sites.

Attributions. The two papers in this chapter are published, the first in *PLoS Pathogens* and the second in *Bioinformatics*. The work is primarily a collaboration between our lab and the talented bioinformatics crew led by Thomas Ioerger at Texas A&M University. With help from Ed Long in Chris Sassetti's lab at the University of Massachusetts Medical School in Worcester, I made the transposon libraries in Mtb and did the DNA preparation for sequencing. The bioinformatic analyses were a joint effort by Chris Sassetti, Curtis Huttenhower in the Biostatistics Department at HSPH, Thomas Ioerger and his student Michael DeJesus, and Jim Sacchettini. I drafted the first manuscript and Michael and Tom drafted the second manuscript, and we were all involved in editing the manuscripts. Since Michael was the primary author on the second manuscript, it is appended in this thesis in its final publication formatting. However, because it is a natural extension of this chapter, it is included here (rather than as an appendix) for narrative continuity.

Section 2.2: Global Assessment of Genomic Regions Required for Growth in *Mycobacterium tuberculosis*

Yanjia J. Zhang[1], Thomas R. Ioerger[2], Curtis Huttenhower[3], Jarukit E. Long[5], Christopher M. Sassetti[5], James C. Sacchettini[4], Eric J. Rubin[1*]

[1]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA

[2]Department of Computer Science, Texas A&M University, College Station, TX

[3]Department of Biostatistics, Harvard School of Public Health, Boston, MA

[4]Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX

[5]Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, MA

*To whom correspondence should be addressed:

Address: 4 Blackfan Circle, HIM 1048, Boston, MA 02115

Phone: 617.432.3335

Abstract

Identifying genomic elements required for viability is central to our understanding of the basic physiology of bacterial pathogens. Recently, the combination of high-density mutagenesis and deep sequencing has allowed for the identification of required and conditionally required genes in many bacteria. Genes, however, make up only a part of the complex genomes of important bacterial pathogens. Here, we use an unbiased analysis to comprehensively identify genomic regions, including genes, domains, and intergenic elements, required for the optimal growth of *Mycobacterium tuberculosis,* a major global health pathogen. We found that several proteins jointly contain both domains required for optimal growth and domains that are dispensable. In addition, many non-coding regions, including regulatory elements and non-coding RNAs, are critical for mycobacterial growth. Our analysis shows that the genetic requirements for growth are more complex than can be appreciated using gene-centric analysis.

Author Summary

The significant rise in drug resistant strains of *Mycobacterium tuberculosis* has highlighted the need for new drug targets. Here, we present a novel method of defining genetic elements required for optimal growth, a key first step for identifying potential drug targets. Similar strategies in other bacterial pathogens have traditionally defined a set of essential protein-coding genes. Bacterial genomes, however, contain many other genetic elements, such as small RNAs and non-coding regulatory sequences. Protein-coding genes themselves also often encode more than one functional element, as in the case of multi-domain genes. Therefore, instead of assessing the quantitative requirement of whole genes, we parsed the genome into comprehensive sets of overlapping windows, unbiased by annotation, and scanned the entire genome for regions required for optimal growth. These required regions include whole genes, as expected; but we also discovered genes that contained both required and non-required domains, as well as non protein-coding RNAs required for optimal growth. By expanding our search for required genetic elements, we show that *Mycobacterium tuberculosis* has a complex genome and discover potential drug targets beyond the more limited set of essential genes.

Introduction

Mutagenesis has long been a powerful tool for understanding the roles of genes and other chromosomal elements. Recently, high-density transposon insertion mutagenesis coupled with deep sequencing has enabled comprehensive identification of the required genes in many important bacterial pathogens [1-6]. Defining the protein-coding genes required for bacterial growth identifies both key biological processes and potential targets for drug development. However, protein-coding genes are not the only genetic elements that code for required functions. In densely packed bacterial genomes, many regulatory regions are required for appropriate expression of genes [7]. Moreover, all organisms produce large numbers of non-coding RNAs that can be important under a variety of growth conditions [8-10]. Gene-oriented analyses also look past cases wherein a single gene encodes several differentially important protein domains.

Here, rather then focusing on genes, we take an unbiased approach to create a comprehensive understanding of genomic requirement in *Mycobacterium tuberculosis* (Mtb). We model the Mtb genome as made up of "functional units", a term that encompasses both genes and other genetic elements, many of which have yet to be annotated. By not limiting our analysis to whole-gene regions, we can find otherwise unidentified functional units while also gaining a more nuanced view of the genes required for mycobacterial growth, including critical domains within proteins and non-protein-coding regions that play important roles.

We find approximately 300 protein-coding genes wherein only portions of the coding sequence are required. These include genes, such as *ppm1* and *fhaA,* where we demonstrate that one domain is required for optimal growth whereas other domains are not. Our unbiased analysis also revealed required genomic elements in regions sitting between protein-coding genes. These include two RNAs, the tmRNA and the RNA component of RNaseP, which are required for optimal growth. In addition, we find a number of other regions that influence viability by uncharacterized mechanisms, but whose effects have previously been overlooked by gene-centric analyses.

Results

Deep sequencing for transposon insertion mapping

To perform a comprehensive assessment of Mtb's genetic requirements for growth, we used two ~100,000-clone Mtb libraries generated through high-density transposon mutagenesis of the H37Rv strain [11]. We generated a library of single-insertion mutants by phage delivery of the Himar1 transposon, which randomly inserts into the genome at sites recognized by the TA dinucleotide (Figure 2.1A). We then created transposon-mapping probes by selectively amplifying and sequencing transposon-genome junctions using an Illumina Genome Analyzer 2. Using genome sequences adjacent to the transposon genomic sequences, we were able to map the insertion site of mutants in the library (Figure 2.1B) and count the reads mapped to each insertion site (insertion count, Table S1).

We reasoned that the insertion count should reflect the number of corresponding mutants in the library. To demonstrate this, we picked twelve individual transposon mutants and added each at a known quantity to a manually constructed library. Insertions were again mapped and counted by deep sequencing, and the insertion counts for each site was compared to the known relative quantities of each mutant in the pool (Figure 2.1C). Insertion counts were highly correlated with the known relative amount of each mutant (Pearson R = 0.880, p-value < 0.0001, n = 14). Additionally, we confirmed that insertion counts accurately reflected the library's genome composition by counting the genome-transposon templates represented in our Illumina reads. Since random shearing events create the genome-transposon templates for amplification, the distance between the transposon and the sheared end represents a unique identifier for each template. We assessed the relationship between estimates of unique template molecules for each TA site and the read count for that site (Figure 2.1D), revealing excellent correlation (Pearson R = 0.945 p-value < 0.0001 n = 36,488).

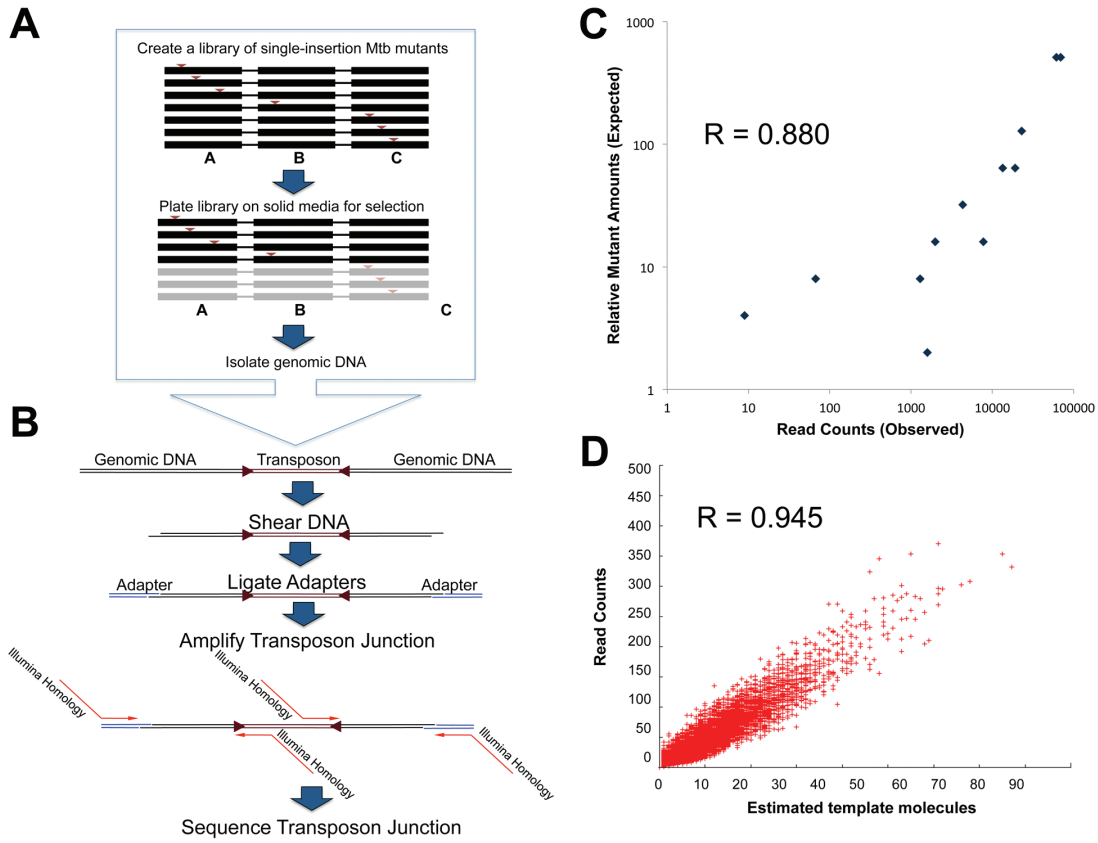


Figure 2.1. Transposon junction sequencing accurately reflects true library content A. A Mtb mutant library is created by phage-delivery of transposons, disrupting each genome with a single insertion. Shown is a schematic of 6 mutant chromosomes spanning three genes (A-C), with transposons—red arrows—disrupting one of the three genes. After growing the library on 7H10 media, we pooled surviving mutants. In this schematic, gene C is required for optimal growth and thus mutants with transposons in gene C are lost. We isolated genomic DNA from the survivors for transposon site mapping. B. We sheared the genomic DNA by sonication, and repaired frayed ends to create blunt ends. We then used Taq polymerase to generate A-tails, allowing the ligation of T-tailed adapters. Finally, we selectively amplified transposon junctions using primers recognizing the transposon end and the adapter. Primers used for amplification contain all requisite sequences to permit direct sequencing of amplicons on an Illumina Genome Analyzer 2. C. We created a library of identified transposon insertion mutants in known relative quantities. DNA from the library was prepared for transposon junction sequencing. Insertion counts were plotted against the known relative quantity of the mutant in the library. D. To further confirm that read counts were a representation of the number of genomes in the library, we estimated the number of PCR template molecules. For each gene, we plotted the estimate of template molecule count against the read counts.

Comprehensive map of genetic requirement in Mtb

In our Mtb library, transposon insertions occurred at 36,488 of the 72,927 possible insertion sites (TA dinucleotides). Each library generated an average of 2.3 million reads, resulting in a mean insertion count of 64 per hit-site. We counted the number of sequencing reads from each site in the two libraries and compiled the counts correcting for each library's total insertion count.

Having demonstrated that insertion counts faithfully represent mutant numbers (Figure 2.1C), we used insertion counts to comprehensively assess the relative importance of selected genomic regions. We defined a region as required for optimal growth if total regional insertions were statistically underrepresented compared to genomic controls (Figure 2.2A). Required regions, therefore, are those in which mutations result in a statistically validated growth defect. We employed a non-parametric test to assess statistical underrepresentation and regions with a p-value of less than 0.01 and a false discovery rate (Benjamini-Hochberg) of less than 0.1 were defined as required for optimal growth *in vitro*.

Instead of assessing the requirement for growth of genomic regions based on predetermined gene coordinates, we divided the genome into contiguous overlapping windows to assess a comprehensive set of potential functional units. Our non-parametric test was powered to find significant regions containing at least 7 TA sites (6 or fewer precluded confident rejection of a null hypothesis of variation by chance alone). Thus, we focused on regions of sizes likely to contain 7 or more TA sites. The mean number of TA sites in windows of 400, 500, and 600 bp was 6.75, 8.45 and 10.12, respectively, and were thus used for our sliding

window analysis of functional requirement for growth. Intergenic (IG) regions are relatively AT-rich in the Mtb genome, allowing us to add a 250 bp sliding window (mean of 6.20 TA sites in IG regions) to the analysis of IG regions. To lower the computational demands of this analysis, we chose to analyze every tenth window, reasoning also that functional units were unlikely to be smaller than 10 base pairs. Thus, we assessed the requirement for growth of every tenth 400, 500, and 600 bp window in the genome, along with every tenth 250 bp window in regions between protein-coding genes (Figure 2.2B).

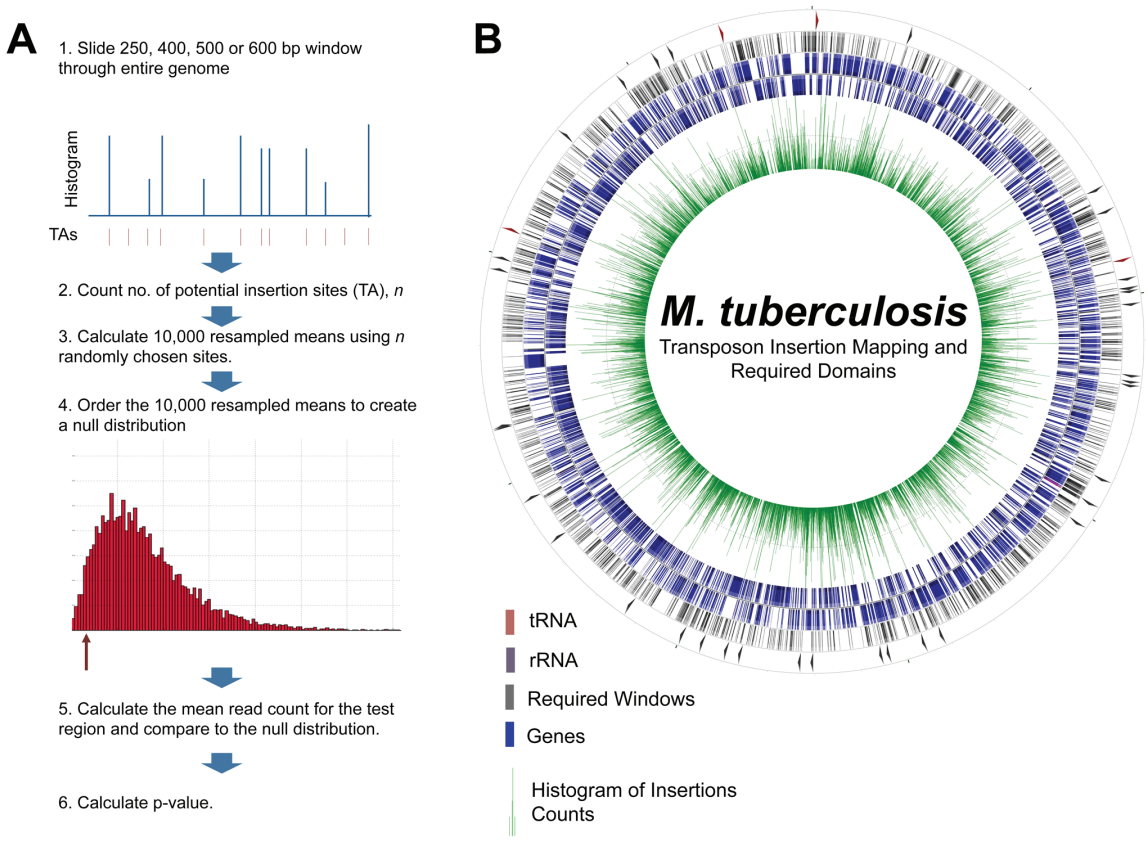


Figure 2.2. Functional requirement testing and mapping
A. Required regions were defined as regions with a statistical underrepresentation of insertion counts compared to the rest of the genome. To test this, we applied a non-parametric test for regions of increasing size, as described in B. Every 250, 400, 500, and 600 bp region (large enough for statistical power) was tested for insertion count underrepresentation, generating a comprehensive map of required regions in the Mtb genome. Tracks on the circularized genome, from inner-most to outer-most: 1. Histogram of insertion counts, 2. Annotated genes, forward direction, 3. Annotated genes, reverse direction, 4. All required regions. 5. Required intergenic regions.

Genes required for optimal growth

We overlaid the coordinates of known genes on the generated results to find those that contained regions required for optimal growth. Of the 3,989 annotated genes, 698 contained required functional units, 3,071 contained no required functional units, while 210 did not sustain insertions but also did not contain enough TAs to meet statistical requirements (Figure 2.3A). As a screen for genes with multiple functional units of varying requirement, we searched for genes that contained both required and non-required regions. A total of 365 genes met these criteria (Figure 2.3A).

Our finding that many genes contain both required and non-required regions suggested that using only whole genes for analysis could misrepresent their importance for growth. Either the entire gene could appear required for optimal growth or the entire gene would be considered dispensable, leaving no room for the possibility that only a segment of the gene might be required. To determine how our results would compare to a gene-centric analysis, we calculated the requirement for growth of each gene by applying the non-parametric test to the gene as a whole. As expected, genes with required segments had a wide range of p-values when assessed using annotated boundaries instead of unbiased overlapping windows (Figure 2.3B, dark blue bars). A total of 151 out of the 365 (41.3%) had p-values above 0.01, demonstrating that our sliding window strategy accounted for a significant number of required functional units that would be ignored by gene-only strategies (Figure 2.3B).

A transposon insertion into the 5' end of a gene will often block production of the encoded protein, either by attenuating transcription or disrupting ribosome binding sites and initiation codons. Thus, we expected that regions required for optimal growth would tend to be found at the 5' ends of predicted genes. Surprisingly, a plot of the likelihood of discovering an required functional region as a function of the intragenic location revealed a symmetric curve, demonstrating that the required regions discovered have an equal likelihood of residing on either end of the gene (Figure 2.3C). We hypothesized that this may be because the transposon contains a promoter that can direct downstream transcription. To test this, we took two strains that contained transposon insertions and measured mRNA expression upstream and downstream of the transposon (Figure 2.4A). In both cases, expression upstream of the transposon did not significantly change, while downstream expression increased (Figure 2.4B). This is consistent with the observation that downstream genes can be transcriptionally activated by transposon insertions [12]. In addition, mycobacteria are able to use several initiation codons thus making it more likely that truncated but functional proteins can be produced from internal start sites.

While we were able to use this analysis to make many novel observations, we also found that our results supported previous findings. The majority of genes (63%) described as fully required for growth were similarly required in microarray-based studies using transposon site hybridization (TraSH) (Figure S1A, Table S2) [13]. The increased resolution from deep sequencing demonstrated that genes with fewer than 7 TAs resulting in an undersampling

that prevented statistically confident requirement assessments (a separate category for genes with 6 or fewer TAs that did not contain insertions is noted in Figure 2.3A). Since this was not known previously, we predicted that the microarray-determined set of required genes would be biased towards small genes. This proved to be true. In genes predicted to be required by TraSH but not in this study, the average number of TAs was 9.90 (Figure 2.5B). In contrast, the average number of TAs in fully required genes from this study was 19.42, a fair representation of the average of all genes assessed (19.47). In fact, of the genes only determined to be required in TraSH and not in this study, 43% had 7 or fewer TAs, accounting for much of the discordance between the two methods.

A more nuanced analysis of Mtb transposon insertion maps defined essential genes as those that contained “gaps,” any statistically significant runs of potential insertion sites lacking transposon insertions [3]. As expected, genes found in our sliding window analysis to have both required and non-required regions were more concordant with essential genes found by sequencing using this gap analysis than with microarray approaches or whole-gene analyses of insertion counts. Of genes described in our approach as fully required, 97.2% were described as “essential” by Griffin et al (Figure 2.5A), a remarkable level of agreement given the differences in growth media between the two studies. The increased concordance extended to genes containing both required and non-required regions. Griffin et al. described 185 of these genes as essential, while microarray methods only deemed 134 to be essential. The search for required regions within genes, a feature of both analyses, allowed for the discovery of

these regions in longer genes, as evidenced by the increase in average number of TAs within these genes (Figure 2.5B).

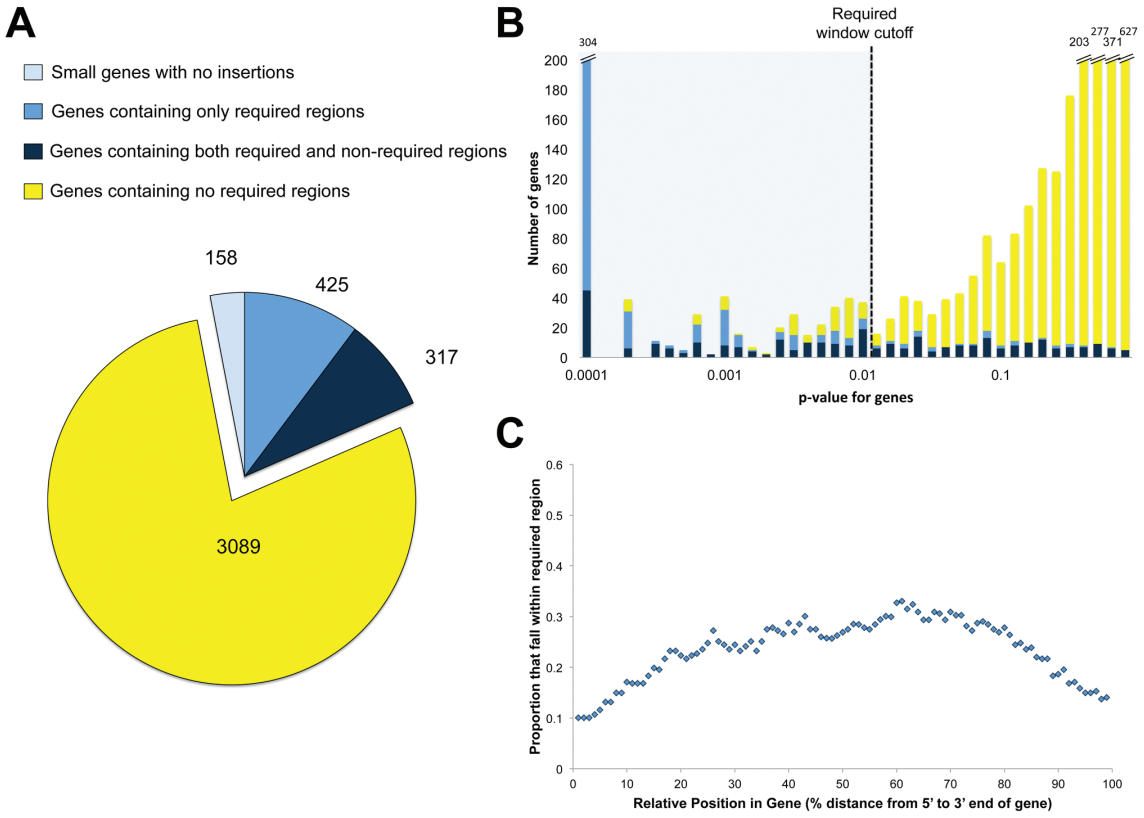


Figure 2.3. Domain Discovery
A. Genes categorized by domain-level resolution of regional requirement. B. Genes categorized as containing only required regions (blue), containing both required and non-required regions (navy) and containing no required regions (yellow) were assessed for requirement along the entire length of the gene, creating a single p-value describing the statistical underrepresentation of insertion reads within the whole gene. For each category, the number of genes across the range of p-values are plotted. C. For genes with both required and non-required regions, the likelihood that the relative position within the gene resides in a required region.

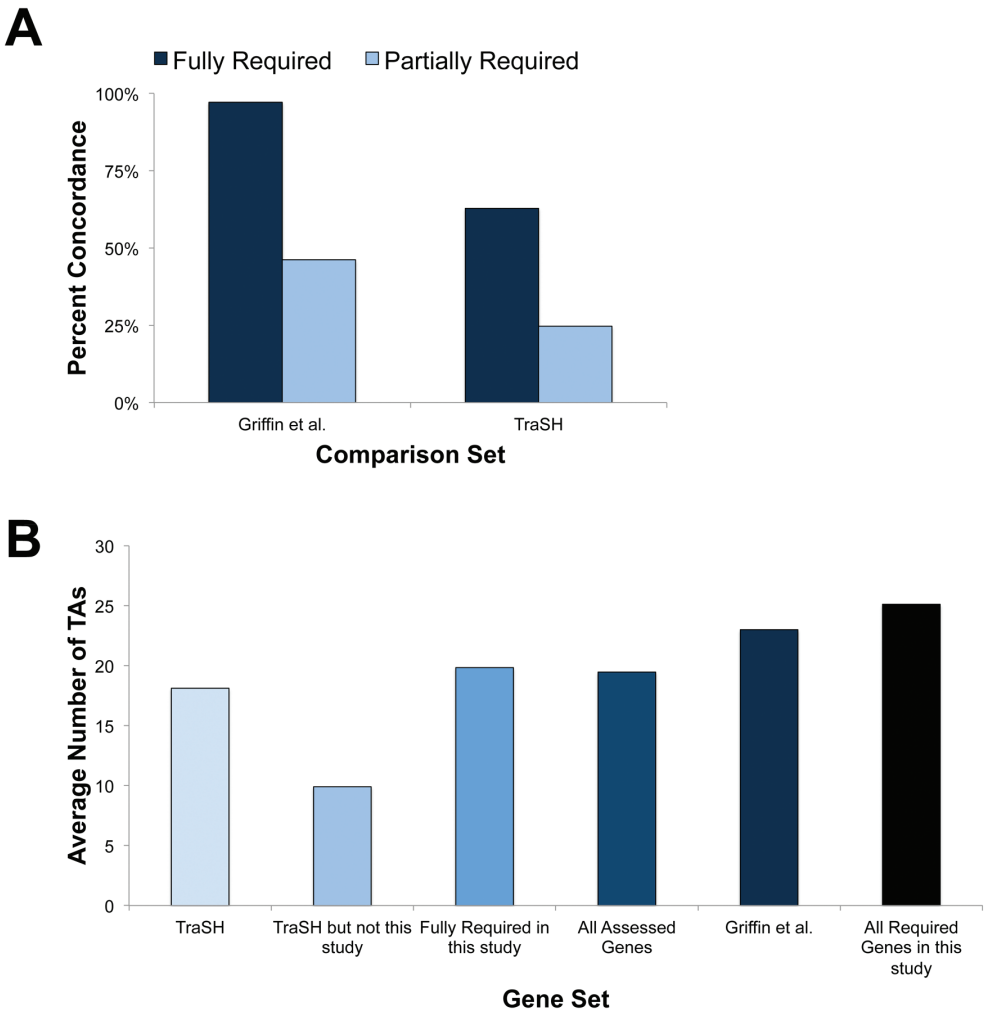


Figure 2.4
A. For both fully required and partially required genes, the agreement with the essential gene set from Griffin et al. and with the required gene set from TraSH was calculated. B. Average length of genes (by number of TAs) are plotted for the following gene sets: required genes in TraSH, genes determined to be required in TraSH but not this analysis, fully required genes in this analysis, all genes assessed in this analysis, essential genes in Griffin et al., and all genes with required regions in this analysis.

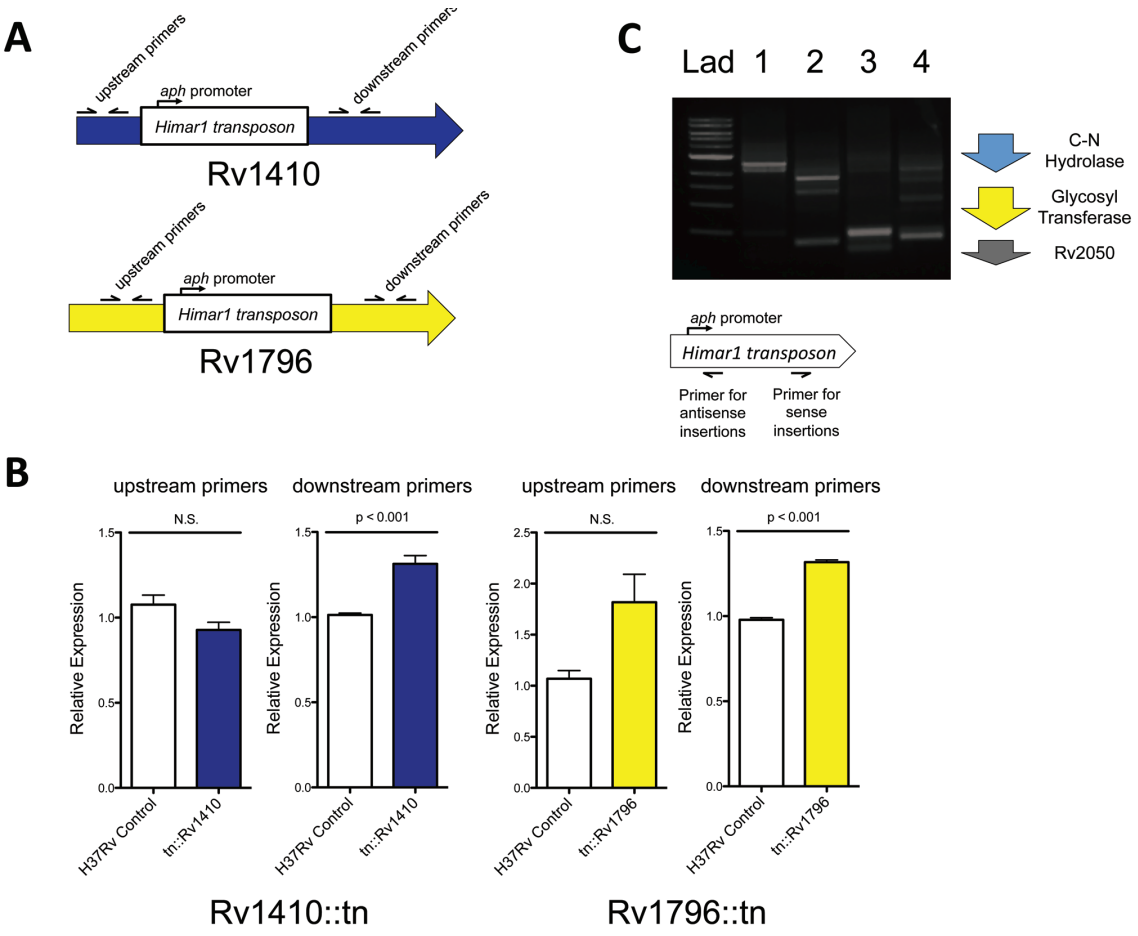


Figure 2.5
A. Primer design for measuring expression upstream and downstream of the transposon. B. Relative expression of gene segments upstream and downstream of the transposon in transposon strains and H37Rv controls. C. Directional footprinting of transposon insertion. Lad: 1 kb DNA ladder, 1: wt *Mtb* transposon library, sense insertions, 2: *ppm1*-complemented Mtb transposon library, sense insertions, 3: wt *Mtb* transposon library, anti-sense insertions, 4: *ppm1*-complemented Mtb transposon library, anti-sense insertions.

Identification of the required glycosyl transferase domain in Ppm1

We find that, in some genes, encoded domains have different effects on growth, accounting for the varying degrees of requirement across the gene's open reading frame. For example, the gene encoding Ppm1, an enzyme in the lipoarabinomannan (LAM) synthesis pathway, encodes a protein with two distinct domains. The region encoding the carbon-nitrogen hydrolase domain of Ppm1 sustained many insertions, while the region encoding the C-terminal glycosyl transferase was required for optimal growth (Figure 2.6A). While the specific requirement of the glycosyl transferase is a novel finding, it resonates with a previous report that only the glycosyl transferase was required for the synthesis of LAM, thought to be an essential cell wall component [14]. Another study revealed that Ppm1 has N-acyltransferase activity, which could be the non-required function of this two-domain protein [15].

To confirm that the lack of insertions in this domain was due to a functional requirement and not to insertional bias or the generation of toxic fusions or truncations, we created transposon libraries in the presence of a second copy of *ppm1*. We reasoned that a second copy would render the endogenous gene non-required and thus permissive for transposon insertion. We designed footprinting PCR primers upstream of the original *ppm1* to specifically generate amplicons containing transposon insertions into the endogenous copy (Figure 2.6B). Footprinting of the original library confirmed our sequencing results, as no insertions were found in the region encoding the glycosyl transferase. However, in the complemented library, that region did contain insertions,

suggesting the glycosyl transferase is functionally required for growth. We further reasoned that only sense insertions—that is, insertions wherein the transposon's internal promoter is oriented in the same direction as the disrupted gene—would be tolerated in the 5' end of *ppm1* to allow for the expression of the C-terminal required domain. To assess this, we used primers specifically designed to amplify sense and anti-sense insertions, and noted only sense insertions in the 5' end (Figure 2.5C). In addition, we confirmed that many in-frame internal start sites exist between 5' transposon insertion sites and the beginning of the 3' domain.

The MviN-binding domain of FhaA is required for growth

A recent report showed that FhaA was required for optimal growth of *Mycobacterium smegmatis* and postulated that the importance of the interaction of FhaA with the essential protein MviN for appropriate regulation of growth and peptidoglycan synthesis [16]. These processes are essential for mycobacterial cell division and cell wall biosynthesis. This work further demonstrated the C-terminal forkhead associated (FHA) domain of FhaA was required for MviN-binding, while an N-terminal domain of unknown function was dispensable for this interaction. In agreement with these findings, we show here that the region of *fhaA* encoding the FHA domain cannot sustain insertions, while the remainder of the gene is dispensable (Figure 2.6C). We used insertion footprinting to confirm these results, and found that the C-terminal insertion mutants were rescued for growth in the presence of a second copy of *fhaA* (Figure 2.6D).

Notably, both *ppm1* and *fhaA*, which we predict to be required for optimal growth based on the presence of a required region within these genes, were classified as non-essential in a previous microarray-based screen [13]. In fact, 231 of the 365 genes containing both required and non-required regions were not previously described as necessary for growth, likely due to the decreased spatial resolution of microarray-based methods. Microarrays limited the resolution of requirement testing to genes, and each gene received a single metric describing its requirement for growth. In addition, as our approach is not confined to gene boundaries, we have the additional resolution to identify domains within genes, as exemplified by *ppm1* and *fhaA*.

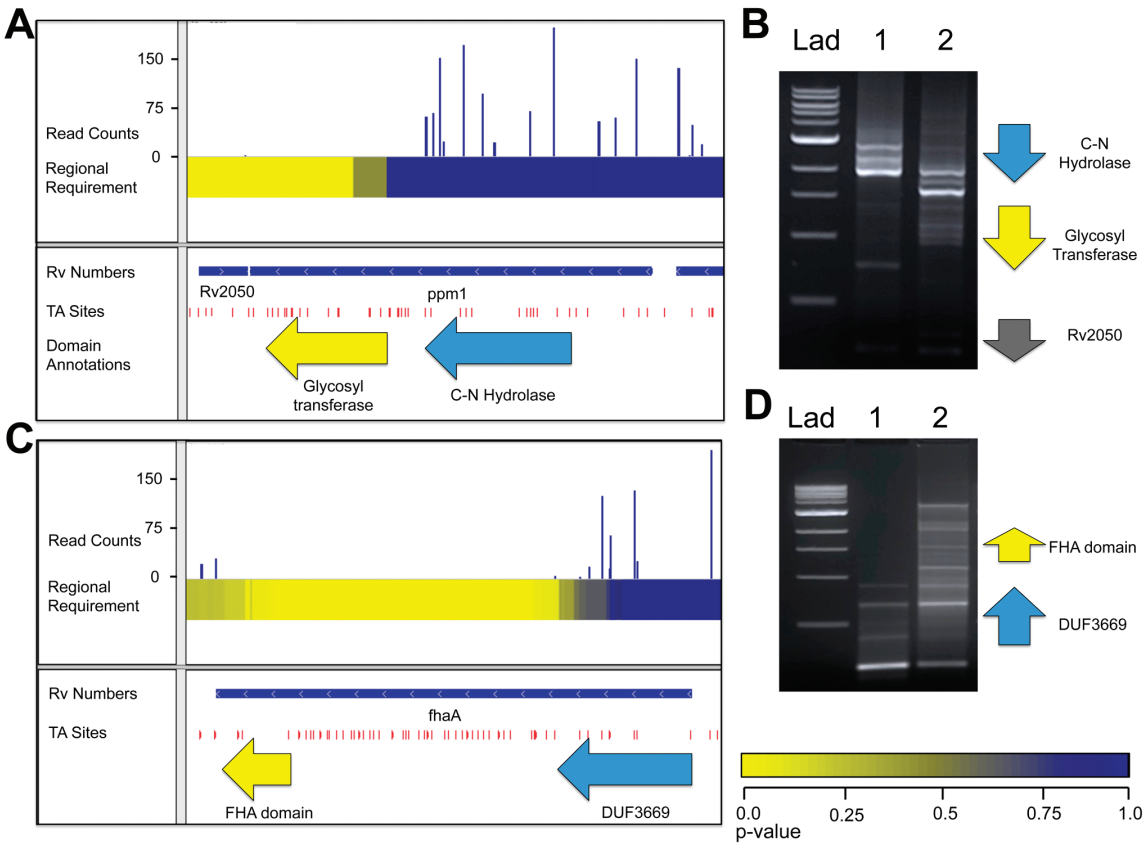


Figure 2.6. *ppm1* and *fhaA* each code for two domains with varying requirements for growth

A. IGV plot for genomic region containing *ppm1*. Tracks, from top to bottom: 1. Histogram of insertion counts, 2. Comprehensive heat-map of requirement of 500-bp windows, 3. Position of annotated genes, 4. TA sites, 5. Position of known domains within *ppm1*. B. PCR footprinting for insertions was performed using primers against the an upstream genomic region and the transposon end, resulting in amplicons spanning *ppm1* to various inserted transposons. Lad: 1 kb DNA ladder, 1: wt Mtb transposon library, 2: *ppm1*-complemented Mtb transposon library. C. IGV plot for genomic region containing *ppm1*. Tracks, from top to bottom: 1. Histogram of insertion counts, 2. Comprehensive heat-map of requirement of 500-bp windows, 3. Position of annotated genes, 4. TA sites. D. PCR footprinting for insertions was performed using primers against the an upstream genomic region and the transposon end, resulting in amplicons spanning *fhaA* to various inserted transposons. Lad: 1 kb DNA ladder, 1: wt Mtb transposon library, 2: *fhaA*-complemented Mtb transposon library.

The RNA components of RNaseP and the Clp quality control system are required for growth

Because we are not limited to annotated regions we were also able to probe the importance of intergenic regions. By scanning the genome for required 250, 400 and 500 bp regions, we found 25 intergenic regions required for optimal growth (Figure 2.2B). These required intergenic regions contained many components of known essential cellular functions to be required for *in vitro* growth. These included 10 tRNAs as well as the RNA catalytic unit of RNaseP, which has been shown to be required for tRNA processing in other bacteria (Figure 2.7A). Additionally, one required intergenic region contained the tmRNA, a molecule required to release stalled ribosomes and to tag polypeptides for proteolytic degradation through an essential protease (Figure 2.7B) [17-18]. Of the intergenic segments containing functionally required regions, 11 had annotated functions and an additional 6 were adjacent to genes assessed as required for growth and, therefore, might contain promoters or other transcriptional regulatory elements. The remaining 19 required segments are situated between two non-required genes and, as yet, have no ascribed function.

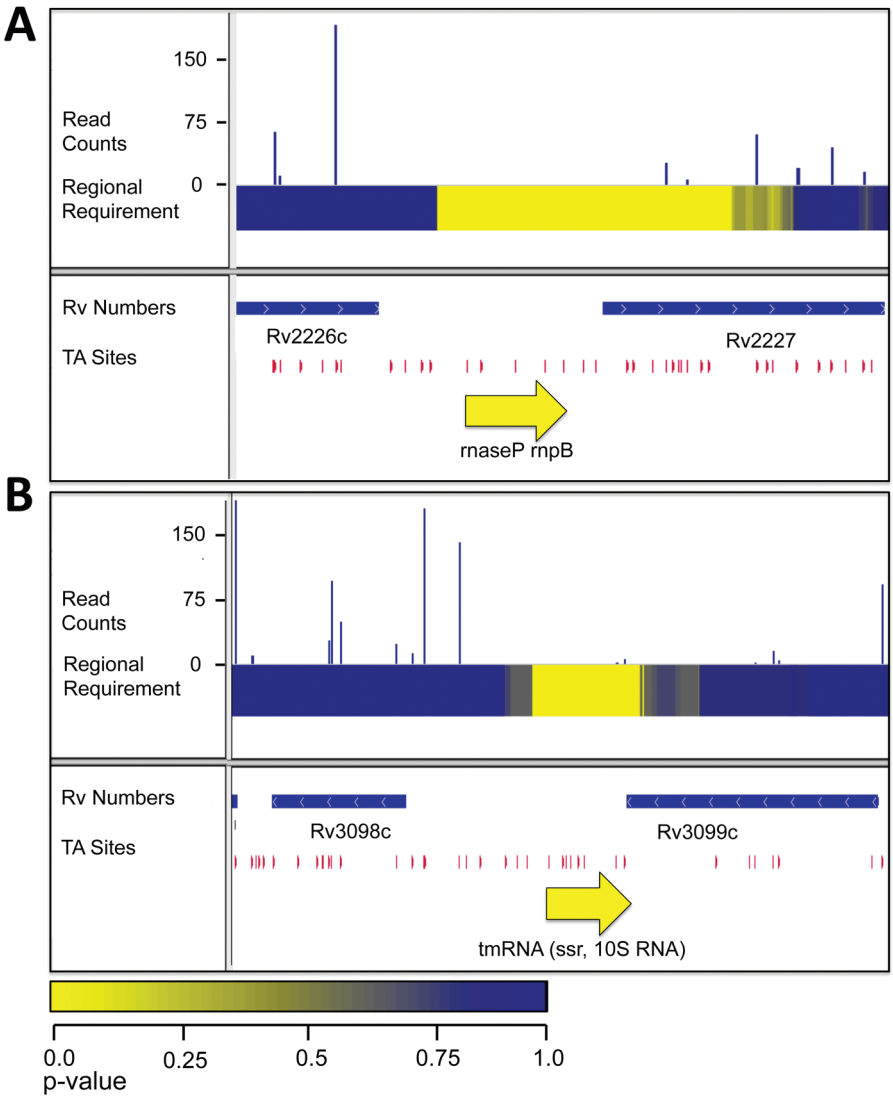


Figure 2.7. RNAs required for growth *in vitro*
A. IGV plot for genomic region containing the rnpB, the RNA component of RNaseL. Tracks, from top to bottom: 1. Histogram of insertion counts, 2. Comprehensive heat-map of requirement of 500-bp windows, 3. Position of annotated genes, 4. Position of TA dinucleotide sites, 5. Position of rnpB. B. IGV plot for genomic region containing the tmRNA. Tracks, from top to bottom: 1. Histogram of insertion counts, 2. Comprehensive heat-map of requirement of 500-bp windows, 3. Position of annotated genes, 4. Position of TA dinucleotide sites, 5. Position of the tmRNA.

Discussion

Finding genetic loci that are required for optimal growth under specific conditions helps inform the basic understanding of bacterial physiology and efforts to develop new therapeutics for pathogens. Previously, we and others have used transposon mutagenesis to infer the requirement for genes under different growth conditions by utilizing the information provided by genome annotations [1-6]. Deep sequencing, which allows us to map precisely the insertion site of every mutant, affords a higher resolution assessment of genetic requirement, beyond just genes. Here, we demonstrate that an unbiased sliding window approach harnesses the full potential of this increased resolution. This approach identified not only whole genes required for optimal growth but also other required elements, such as non-protein coding RNAs and protein domains within insertion-containing genes, which would otherwise obscured by gene-centric analysis. An alternative analysis that uses significant gaps in insertion—rather than quantitative insertion counts—was also able to assess the requirement of protein domains (DeJesus et al., unpublished data, submitted). This analysis likely identifies regions absolutely essential for viability rather than all regions required for optimal growth.

We found that many genes contain elements that are important for growth even though other regions are not required. In at least two cases, *ppm1* and *fhaA*, published data have shown that the required regions encode specific protein domains. However, in other cases, these might represent non-protein-coding RNAs or cis regulatory elements. Bacteria encode many small RNAs many of which could be required for optimal growth and some of which are

embedded within genes [8-10]. In addition, most genes have been annotated computationally, an uncertain pursuit that clearly can lead to misannotated start sites [19]. Genes with only 5' insertions could fall into this category.

Similarly, important non-protein-coding regions could have multiple roles. In some cases, we found that known RNAs, such as *rnpB*, the catalytic RNA component of RNase P, and the tmRNA were required for optimal growth, supporting previous speculation [20-21]. Again, some other required regions might encode as yet unidentified non-coding RNA molecules. Still others might be promoters or other regulatory regions.

In this study, our resolution was limited by the specific properties of the *Himar1* transposon in mycobacteria. Our previous studies have shown that insertions are randomly distributed apart from the desired selection against insertion in essential regions [11, 22]. Despite this, we cannot assume that all sites lacking insertions represent required regions since unknown insertional biases of the transposon may exist. Thus, we defined a required region as one with a statistically underrepresented insertion count using a non-parametric test to account for such potentially unique biases within these data (Figure 2A). This allowed us to exclude, for example, windows with 6 or fewer TA sites, which demonstrably lacked power to distinguish a region as essential for growth relative to background variation. In GC-rich protein-coding regions, this limited our scope to windows of greater than 400 bp; less GC-rich intergenic regions allowed the assessment of windows greater than 250 bp. Thus, while we were able to identify many required protein domains and RNAs, it is certainly possible that smaller

elements required for growth were missed due to these size constraints. This is a particular problem for non-coding RNAs that are often very small. For example, while we found 10 tRNAs required for growth, the remaining tRNAs reside in non-coding regions that did not have the requisite number of TA sites to determine requirement. Using the *Himar1* transposon in organisms with less of a GC bias, or in organisms in which a less restricted transposon exists, should result in increased resolution [4].

The analysis we used provides a powerful tool to perform functional genome analysis. Importantly, this type of approach is useful not only for single conditions, as we described but can also be used to identify elements critical under one growth condition but not another [23-25]. This is particularly important in organisms like Mtb, an obligate pathogen that never grows under conditions precisely comparable to those we use *in vitro*. Coupling high-density insertion libraries with deep sequencing and analytic methods such as that described here provides a powerful experimental tool for functional genome annotation.

Methods and Materials

Genomic Library Creation

Two independent libraries of 100,000 mutants were generated in the Mtb strain H37Rv as previously described on 7H10 agar [11]. Independent libraries were also generated in Mtb strains overexpressing *ppm1* and *fhaA*. Genomic DNA was isolated from each library and randomly fragmented to 400-600 bp pieces by sonication with a Covaris E220. Nicked ends were repaired (Epicentre end repaired kit), and A-tails were added with Taq polymerase to allow the

ligation of T-tailed adapters. Transposon-junctions were amplified for 30 cycles (94 degrees, 30 seconds; 58 degrees, 30 seconds; 72 degrees, 30 seconds) using a primer recognizing the transposon end (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCCGGGGACTTATCAGCCAACC-3') and one recognizing the adapter (5'-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCGTCCAGTCTCGCAGATGATAAGG-3'). Primers used during amplification contained all the requisite sequence for binding to the Illumina sequencing platform. A 250-400 bp fragment of the amplicon was isolated from a gel and sequenced on an Illumina GA2 instrument with a custom sequencing primer (5'-TTCCGATCCGGGGACTTATCAGCCAACC-3').

Sequencing analysis

Reads from the Illumina sequencing run were first screened for the presence of sequence from the end of the Himar1 transposon. The following 35 bases were mapped to the Mtb genome, allowing for 2 mismatches. Reads that mapped to the genome at a TA site were designated as mapped insertions. Reads that mapped to multiple sites were randomly assigned to one of the mapped sites. For each library, the number of reads mapping to each site (insertion counts) was counted. Insertion counts were plotted on IGV and CGViewer [26-27].

Requirement Testing

For every possible region size containing *x* potential insertion sites, a null distribution of mean read counts was generated by calculating the mean read

counts from a set of 10,000 randomly selected sets of *x* sites. The 10,000 randomly generated means were sorted and the rank of the test region's mean insertion count within the ordered null distribution was determined. The p-value was calculated as the rank of the test mean divided by the size the null distribution (10,000). Multiple test correction was performed by calculating the Benjamini-Hochberg false discovery rate over all regions tested. Regions containing 7 TA sites with no insertions had a p-value of 0.008 and an FDR of 0.06, while regions containing 6 TA sites with no insertions had a p-value of 0.018 and an FDR of 0.12. In order to power our study to detect required regions containing at least 7 TAs, we determined a region to be required for optimal growth if it had a p-value less than 0.01 and an FDR less than 0.1.

DNA footprinting

Footprinting of transposon insertion sites was performed by PCR using a primer recognizing the Himar1 ITR sequence (5'-CCCGAAAAGTGCCACCTAAATTGTAAGCG-3') and primers recognizing a genomic segment just upstream of the gene of interest. For directional footprinting, we used one primer to amplify sense insertions (5'-TTTTCTGGATTCATCGACTGTGGC-3')—where the kanamycin resistance gene on the transposon was oriented in the same direction as the disrupted gene—and another for antisense insertions (5'-CAGCTCATTTTTTAACCAATAGGCCG-3'). Standard PCR conditions were used for long amplification with Phusion polymerase (94 degrees, 15 seconds; primer-dependent annealing temperature, 30 seconds; 72 degrees, 2 minutes).

Acknowledgements

We thank Jennifer Griffin and Michael DeJesus for their insightful comments and for their continued collaboration, Amanda Martinot for her transposon insertion strains and thoughtful discussions, Jim Robinson for his tremendous help with IGV plotting, and Paul Stothard for his aid in genomic data representation on CGViewer.

Section 2.3: Bibliography

1. Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. Proc Natl Acad Sci USA 106: 16422–16427. doi:10.1073/pnas.0906627106.

2. Goodman AL, Mcnulty NP, Zhao Y, Leip D, Mitra RD, et al. (2009) Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. Cell Host Microbe 6: 279–289. doi:10.1016/j.chom.2009.08.003.

3. Griffin JE, Gawronski JD, DeJesus MA, Ioerger TR, Akerley BJ, et al. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog 7: e1002251. doi:10.1371/journal.ppat.1002251.

4. Langridge GC, Phan M-D, Turner DJ, Perkins TT, Parts L, et al. (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Research 19: 2308–2316. doi:10.1101/gr.097097.109.

5. van Opijnen T, Bodi KL, Camilli A (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods 6: 767–772. doi:10.1038/nmeth.1377.

6. Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Ben Passarelli, et al. (2011) The essential genome of a bacterium. Mol Syst Biol 7: 1–7. doi:10.1038/msb.2011.58.

7. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, et al. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393: 537–544. doi:10.1038/31159.

8. Arnvig KB, Young DB (2009) Identification of small RNAs in Mycobacterium tuberculosis. Mol Microbiol 73: 397–408. doi:10.1111/j.1365-2958.2009.06777.x.

9. DiChiara JM, Contreras-Martinez LM, Livny J, Smith D, McDonough KA, et al. (2010) Multiple small RNAs identified in Mycobacterium bovis BCG are also expressed in Mycobacterium tuberculosis and Mycobacterium smegmatis. Nucleic Acids Research 38: 4067–4078. doi:10.1093/nar/gkq101.

10. Arnvig KB, Comas I, Thomson NR, Houghton J, Boshoff HI, et al. (2011) Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis. PLoS Pathog 7: e1002342.

doi:10.1371/journal.ppat.1002342.

11. Sassetti CM, Boyd DH, Rubin EJ (2001) Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci USA 98: 12712–12717. doi:10.1073/pnas.231275498.

12. Rubin EJ, Akerley BJ, Novik VN, Lampe DJ, Husson RN, et al. (1999) In vivo transposition of mariner-based elements in enteric bacteria and mycobacteria. Proc Natl Acad Sci USA 96: 1645–1650.
13. Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 48: 77–84.

14. Gurcha SS, Baulard AR, Kremer L, Locht C, Moody DB, et al. (2002) Ppm1, a novel polyprenol monophosphomannose synthase from Mycobacterium tuberculosis. Biochem J 365: 441–450. doi:10.1042/BJ20020107.

15. Tschumi A, Nai C, Auchli Y, Hunziker P, Gehrig P, et al. (2009) Identification of apolipoprotein N-acyltransferase (Lnt) in mycobacteria. Journal of Biological Chemistry 284: 27146–27156. doi:10.1074/jbc.M109.022715.

16. Gee CL, Papavinasasundaram KG, Blair SR, Baer CE, Falick AM, et al. (2012) A phosphorylated pseudokinase complex controls cell wall synthesis in mycobacteria. Sci Signal 5: ra7. doi:10.1126/scisignal.2002525.

17. Tu GF, Reid GE, Zhang JG, Moritz RL, Simpson RJ (1995) C-terminal extension of truncated recombinant proteins in Escherichia coli with a 10Sa RNA decapeptide. J Biol Chem 270: 9322–9326.

18. Gottesman S, Roche E, Zhou Y, Sauer RT (1998) The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. Genes Dev 12: 1338–1347.

19. Rosenkrands I, Weldingh K, Jacobsen S, Hansen CV, Florio W, et al. (2000) Mapping and identification of Mycobacterium tuberculosis proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection. Electrophoresis 21: 935–948. doi:10.1002/(SICI)1522-2683(20000301)21:5<935::AID-ELPS935>3.0.CO;2-P.

20. Stark BC, Kole R, Bowman EJ, Altman S (1978) Ribonuclease P: an enzyme with an essential RNA component. Proc Natl Acad Sci USA 75: 3717–3721.

21. Kole R, Baer MF, Stark BC, Altman S (1980) E. coli RNAase P has a required RNA component. Cell 19: 881–887.

22. Rubin EJ, Akerley BJ, Novik VN, Lampe DJ, Husson RN, et al. (1999) In vivo transposition of mariner-based elements in enteric bacteria and mycobacteria. Proc Natl Acad Sci USA 96: 1645–1650.

23. Sassetti CM, Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci USA 100: 12989–12994. doi:10.1073/pnas.2134250100.

24. Rengarajan J, Bloom BR, Rubin EJ (2005) Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. Proc Natl Acad Sci USA 102: 8327–8332. doi:10.1073/pnas.0503272102.

25. Murry JP, Pandey AK, Sassetti CM, Rubin EJ (2009) Phthiocerol dimycocerosate transport is required for resisting interferon-gamma-independent immunity. J INFECT DIS 200: 774–782. doi:10.1086/605128.

26. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, et al. (2011) Integrative genomics viewer. Nat Biotechnol 29: 24–26. doi:10.1038/nbt0111-24.

27. Grant JR, Stothard P (2008) The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Research 36: W181–W184. doi:10.1093/nar/gkn179.

Section 2.4: Bayesian Analysis of Gene Essentiality based on Sequencing of Transposon Insertion Libraries

Michael DeJesus1, Yanjia J. Zhang2, Christopher M. Sassetti3, Eric J. Rubin2, James C. Sacchettini4, and Thomas R. Ioerger1*

1Department of Computer Science, Texas A&M University, College Station, TX 2Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 3Department of Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, MA 4Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX

Bioinformatics Advance Access published January 29, 2013

Bayesian Analysis of Gene Essentiality based on Sequencing of Transposon Insertion Libraries

Michael DeJesus[1], Yanjia J. Zhang[2], Christopher M. Sassetti[3], Eric J. Rubin[2], James C. Sacchettini[4], and Thomas R. Ioerger[1*]

[1]Department of Computer Science, Texas A&M University, College Station, TX
[2]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA
[3]Department of Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, MA
[4]Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX

Associate Editor: Prof. Alfonso Valencia

ABSTRACT

Motivation: Next-generation sequencing affords an efficient analysis of transposon insertion libraries, which can be used to identify essential genes in bacteria. To analyze this high-resolution data, we present a formal Bayesian framework for estimating the posterior probability of essentiality for each gene, using the Extreme Value distribution to characterize the statistical significance of the longest region lacking insertions within a gene. We describe a sampling procedure based on the Metropolis-Hastings algorithm to calculate posterior probabilities of essentiality while simultaneously integrating over unknown internal parameters.

Results: Using a sequence dataset from a transposon library for *M. tuberculosis*, we show that this Bayesian approach predicts essential genes that correspond well with genes shown to be essential in previous studies. Furthermore, we show that by using the Extreme Value Distribution to characterize genomic regions lacking transposon insertions, this method is capable of identifying essential domains within genes. This approach can be used for analyzing transposon libraries in other organisms, and augmenting essentiality predictions with statistical confidence scores.

Availability: A python script implementing the method described is available for download from http://saclab.tamu.edu/essentiality/

Contact: michael.dejesus@tamu.edu; ioerger@cs.tamu.edu;

Supplementary Information: Supplementary data is available at Bioinformatics online.

1 INTRODUCTION

Transposon mutagenesis is a frequently used laboratory method for determining essential genes in bacterial organisms. Essential genes are those genes necessary for growth under a wide variety of environmental conditions. Knowledge of essential genes is important for the discovery of new antibacterial drugs because these genes are potential targets for inhibitors (Hasan *et al.*, 2006). One way of determining essential genes is to identify regions of the genome in which insertional mutations produce nonviable cells. To do this, a high-density library of transposon mutants is constructed. The synthetic transposons used in these studies are small fragments of DNA (typically 1-2 kb), which can be inserted into different locations in the chromosome through the action of a distally-encoded transposase enzyme (Hayes, 2003). For example, derivatives of the *Himar1* transposon are widely used and have been characterized to insert at arbitrary TA dinucleotides without any other obvious sequence specificity bias (Lampe *et al.*, 1996; Rubin *et al.*, 1999). The total number of TA sites within a gene often varies around 10-100 sites depending on gene length and GC content. When a transposon inserts at one of these TA sites within a gene, it presumably disrupts the function of the gene. In a large library of transposon insertion mutants, genes harboring insertions are presumed to be non-essential. Genes lacking insertions may be essential as they cannot tolerate disruption, however this depends on the size of gene and degree of saturation of the library (Lamichhane *et al.*, 2003). Typically 10-15% of ORFs in a bacterial genome are found to be essential (Gerdes *et al.*, 2003), including genes involved in core metabolism, cell-wall biosynthesis, protein translation, and DNA replication (all of which are known targets of existing drugs). Differential analysis of essential genes in bacteria passaged through a host could be used to identify genes specifically required for infection (Sassetti and Rubin, 2003).

In the original implementation, the location of transposon insertions in individual mutants were read out via microarray hybridization. A primer-extension step, using a primer complimentary to one end of the transposon, was used to amplify the adjacent genomic region, and the relative abundance of these nucleic acid probes was quantified via hybridization to oligonucleotide representing each gene (Sassetti *et al.*, 2003). Both the resolution and the quantitative accuracy of this method were limited. More recently, use of hybridization to analyze transposon libraries has been replaced by deep sequencing using next-generation sequencers, which yield millions of short reads (typically 50-100 bp). Mapping of reads amplified from transposon boundaries can give precise coordinates of insertions within the genome (Gawronski *et al.*, 2009; Langridge *et al.*, 2009; Griffin *et al.*, 2011). The high-resolution data afforded by deep sequencing presents some unique challenges for analysis of

*to whom correspondence should be addressed

gene essentiality. It has previously been observed that even essential genes can tolerate transposon insertions in the extreme N- and C-termini of the ORF (Smith *et al.*, 1996; Akerley *et al.*, 1998; Christen *et al.*, 2011)). Previous analyses have often used an *ad hoc* criterion, such as exclusion of insertions in the first/last 5-20% of the coding region (Gawronski *et al.*, 2009). For similar reasons, insertions are sometimes tolerated in linker regions between domains, or one domain but not another of an essential protein (Lamichhane *et al.*, 2005). For example, transposon insertions in the N-terminus of Mmpl5 caused attenuation of infection in mouse lungs, whereas insertions in the C-terminus did not (Lamichhane *et al.*, 2005). Thus it is inaccurate to assume that only genes completely lacking transposon insertions are essential.

In previous work, we described a novel statistical method for analyzing transposon insertion data to characterize the essentiality of genes within an organism (Griffin *et al.*, 2011). The method was based on identifying the longest consecutive stretch of TA sites lacking insertions in a gene, and estimating the likelihood of such an open region occurring by chance through the Extreme Value (Gumbel) distribution. This model was based on an analogy to runs of tails in a sequence of coin tosses, where each TA site is viewed as an independent Bernoulli trial given the background insertion frequency in non-essential genes. This analysis was shown to correlate well with previous characterizations of genes essential for *in vitro* growth of *M. tuberculosis*. The primary advantage of this method is that essentiality is based on statistically significant stretches of TA sites lacking insertions, regardless of the presence of insertions at other regions within the gene. This is in contrast to other models, such as a multinomial model, where the order of insertions is not taken into consideration (Blades and Broman, 2002), and may miss regions characteristic of essential domains. One limitation of our previous method is that it depends on an *a priori* estimate of the insertion frequency in non-essential genes. While this can be approximated (over all TA sites in the whole genome) or tuned iteratively (by separating out essential genes), a more rigorous statistical treatment is desirable.

One possible way to approximate the parameters of this model and find estimates of essentiality is to use the Expectation Maximization (EM) algorithm. Although the EM algorithm converges relatively quickly, it depends on maximizing the likelihood of the given distribution. This is not feasible for the product of Gumbel distributions, as no closed-form expression for the derivative exists.

In this paper, we present a formal Bayesian analysis of transposon insertion data that simultaneously estimates the likelihood of essentiality for each gene and the non-insertion frequency for each class, given the data. We develop a formula for the joint and conditional densities based on the likelihood for each gene. We describe how to use a Metropolis Hastings sampling procedure to estimate the parameters from the data, by sampling from the joint probability densities. This method produces a formal estimate of essentiality for each gene from the posterior probability given the observed insertion data, marginalizing over the unknown insertion frequencies in essential and non-essential genes.

2 METHODS

The sequence data obtained from transposon mutagenesis experiments consists of a set of reads mapping to TA dinucleotides sites within the genome. The read counts at each TA site are discretized to a binary value of 1 ("insertion") or 0 ("non-insertion") depending on the presence or absence of transposon reads mapping to those locations. While the number of reads mapping to a location may contain useful additional information about essentiality, it can also be subject to variability due to phenomena such as PCR bias. Thus we take the presence/absence of insertions within genes to be sufficient for identifying essential regions in our model. TA sites are assumed to be independent from each other and treated as a set of Bernoulli trials (analogous to coin-tossing, with insertions and non-insertions representing outcomes of Tails or Heads), which is a reasonable assumption in non-essential genes, where the probability of insertion at adjacent sites is thought to be independent.

From this data we obtain the maximum number of consecutive TA sites lacking insertions within the genes and the number of nucleotides spanned by this sequence. While the geometric distribution governs the distribution of the number of non-insertions observed in a row, the Gumbel (Extreme Value Distribution) can be used to characterize the longest run of non-insertions observed in a gene. The Gumbel distribution serves as a likelihood function for non-essential genes, as the longest runs of non-insertion should follow what we expect given the global non-insertion frequency. On the other hand essential genes, whose maximum runs of non-insertions should be longer than expected, are instead modeled through a normalized sigmoid function. This likelihood function reflects the fact that any gene can be essential (with more or less uniform probability), except those genes with spans of non-insertion that are too small to represent a domain. Using this model of the data we derive posterior densities for the essentiality of each gene and use the Metropolis-Hastings algorithm to obtain a MCMC sample of values of these densities from which to estimate their posterior probabilities.

2.1 Bayesian Mixture Model

Let $Y_i = \{n_i, r_i, s_i\}$ represent our observations for the i-th gene for $i = 1...G$, where n_i represents the total number of TA sites, r_i represent the longest run of non-insertions observed, and s_i represents the span of nucleotides of the longest run of non-insertions. Each gene is modeled as coming from one of two classes, 1 and 0, representing essential and non-essential genes. The complete set of essentiality assignments is represented by the latent variable $Z = \langle Z_1, Z_2, ...Z_G \rangle$ (Boolean vector), with the essentiality assignment of an individual i-th gene represented by the boolean variable Z_i which takes on binary values of 1 and 0 for the two possible classes. We assume a Bernoulli probability, ϕ_0, that governs probability of non-insertion across non-essential genes. Finally, $\omega = \langle \omega_1, \omega_0 \rangle$, the mixing coefficient, represents the prevalence of essential and non-essential genes within the mixture (with $\omega_0 = 1 - \omega_1$).

2.1.1 Likelihood for Non-Essential Genes The data, Y_i, for each gene consists of observations r_i and s_i, representing the maximum run of non-insertions (TA sites without insertions) in a row, and the number of nucleotides spanned by this gap of non-insertions. The joint likelihood of these observations is:

$$p(r_i, s_i | Z_i = 0, \phi_0, \omega_1) = p(r_i | Z = 0, \phi_0, \omega_1) \times p(s_i | r_i, Z = 0, \phi_0, \omega_1)$$

To The likelihood of observing a maximum run of non-insertions, r_i, is modeled through the Gumbel distribution:

$$p(r_i|Z_i = 0, \phi_0, \omega_1) = \text{Gumbel}(r_i; m, \tau) = \frac{1}{\tau} e^{-z-e^{-z}} \quad (1)$$

where $z = \frac{r_i - m}{\tau}$, and m and τ are the location and scale parameters of the underlying distribution. In analogy to coin-tossing, these parameters are functions of the probability of non-insertion, ϕ_0, and of the total number of trials, n, derived by determining the expected maximum value in a series of independent samples from a geometric distribution (Schilling, 1990):

$$m = \log_{\frac{1}{\phi_0}}(n(1-\phi_0)) \qquad \tau = \frac{1}{\log \frac{1}{\phi_0}} \quad (2)$$

Note that the expected length of the maximum run of non-insertions scales up logarithmically with n (the total number of TA sites in the gene).

Since r_i and s_i are highly correlated (see Figure 1), we model their dependence as linear-Gaussian, with covariance matrix $\Sigma = \left[\left[\sigma_r^2, \sigma_{r,s}\right], \left[\sigma_{r,s}, \sigma_s^2\right]\right]$ estimated *a priori* from empirical data:

$$p(s_i|r_i, Z = 0, \phi_0, \omega_1) \sim \text{N}(s_i - \lambda_r r_i, \sigma_r^2) \quad (3)$$

were λ_r and σ_r are the parameters of the Normal distribution, derived from the Linear Gaussian relationship (i.e. $\lambda_r = \frac{\sigma_{r,s}}{\sigma_r}$) observed in the data.

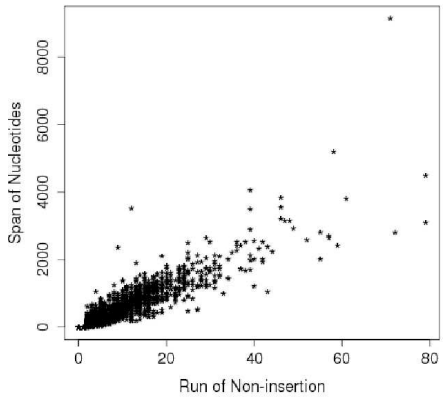


Fig. 1. Relationship Between Length of Run of Non-Insertions (# TA Sites) and Span (nucleotides).

The joint likelihood of the observations at non-essential genes is therefore:

$$p(r_i, s_i|Z_i = 0, \phi_0, \omega_1) = \text{Gumbel}(r_i; m, \tau) \times \text{N}(s_i - \lambda_r r_i, \sigma_r^2) \quad (4)$$

2.1.2 Likelihood for Essential Genes We model the likelihood at essential genes based on a sigmoid function that is uniform as long as the gene contains a gap that is as large as a typical protein domain, since a gap could be any size, even as large as the entire ORF. Using this likelihood allows our method to disambiguate those cases where the run of non-insertions actually represents a smaller or larger segment of the genome than suggested by the number of consecutive TA sites without insertions.

$$p(r_i, s_i|Z_i = 1, \phi_0, \omega_1) = p(s_i|Z = 1, \phi_0, \omega_1) \times p(r_i|s_i, Z = 1, \phi_0, \omega_1)$$

The number of nucleotides spanned by a given run on non-insertions, s_i, is modeled by a normalized sigmoid (logistic) function:

$$p(s_i|Z_i = 1) = \Omega(s_i; \delta) = \frac{C}{1 + e^{\kappa * (\delta - s_i)}} \quad (5)$$

where δ is the mean number of nucleotides spanned by an average protein domain, κ is equal to 0.1, and C is a normalization constant. Previous studies of the length of domains within proteins have found the average size to be roughly 100 amino-acids or 300 bp (Wheelan *et al.*, 2000). Using this threshold for δ, the likelihood of observing a given span s_i is more or less uniform, except it is near 0 if the the longest run of non-insertions spans less than about 300 bp. Figure S2 in the Supplementary Material shows how sensitive the model is to the these parameters.

As with non-essential genes, the likelihood of observing a span of nucleotides r_i given s_i is modeled through a linear-Gaussian dependence similar to Equation (3), but with an inverse relationship (i.e. $N(r_i - \lambda_s s_i, \sigma_s^2)$). The joint likelihood of the observations at essential genes is therefore:

$$p(r_i, s_i|Z_i = 1, \phi_0, \omega_1) = \Omega(s_i) \times \text{N}(r_i - \lambda_s s_i, \sigma_s^2) \quad (6)$$

2.1.3 Prior Distributions Our prior expectation for the probability of non-insertion at non-essential genes, ϕ_0, is represented by a Beta distribution, with hyper-parameters α_0 and β_0:

$$\pi(\phi_0) = \text{Beta}(\phi_0; \alpha_0, \beta_0) = \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \phi_0^{\alpha_0 - 1}(1 - \phi_0)^{\beta_0 - 1}$$

The prior probability of a individual essentiality assignment, Z_i, depends on the probability that the i-th gene is essential or non-essential, and therefore is characterized by a Bernoulli distribution with that depends on ω_1:

$$\pi(Z_i \mid \omega_1) = \text{Bernoulli}(Z_i; \omega_1) = \omega_1^{Z_i}(1 - \omega_1)^{1 - Z_i}$$

Similarly, the prior probability of an essentiality assignment for all genes, Z is a product of Bernoulli trials with probability ω_1:

$$\pi(Z \mid \omega_1) = \prod_i^G \text{Bernoulli}(\omega_1) = \omega_1^{K_z}(1 - \omega_1)^{G - K_z}$$

where G is the total number of genes, and K_z is the sum of the binary vector of essentiality assignments (i.e., $K_z = \Sigma Z_i$). Finally, our prior expectations for the mixing coefficient ω_1 are given by a Beta distribution:

$$\pi(\omega_1) = \text{Beta}(\omega_1; \alpha_w, \beta_w) = \frac{\Gamma(\alpha_w + \beta_w)}{\Gamma(\alpha_w)\Gamma(\beta_w)} \omega_1^{\alpha_w - 1}(1 - \omega_1)^{\beta_w - 1}$$

2.1.4 Joint Distribution To derive the posterior probability density functions necessary for our Bayesian inferences of essentiality, we first define the full joint distribution, $p(Z, Y, \phi_0, \omega_1)$. The full joint distribution is equal to the product of the data-likelihood and the prior expectations for the variables: $p(Z, Y, \phi_0, \omega_1) = p(Y|Z, \phi_0, \omega_1)\ p(\phi_0)\ p(Z|\omega_1)\ p(\omega_1)$. We assume independence among genes, therefore the likelihood can be written as a product of the observations over individual genes:

$$\begin{aligned} p(Y, Z, \phi_0, \omega_1) &= p(Y \mid Z, \phi_0, \omega_1) \times \pi(\phi_0) \times \pi(Z \mid \omega_1) \times \pi(\omega_1) \\ &= \left[\prod_{i=1}^{non} \text{Gumbel}(r_i|\mu, \sigma) \times \text{N}(s_i - \lambda_r r_i, \sigma_r^2)\right] \\ &\times \left[\prod_{i=1}^{ess} \Omega(s_i) \times \text{N}(r_i - \lambda_s s_i, \sigma_s^2)\right] \times \text{Beta}(\phi_0; \alpha_0, \beta_0) \\ &\times \text{Binomial}(K_z; G, \omega_1) \times \text{Beta}(\omega_1; \alpha_w, \beta_w) \end{aligned} \tag{7}$$

2.1.5 Conditional Distributions Using the full joint probability (7) we derive a conditional distribution for the probability of non-insertion at non-essential genes, ϕ_0, using proportionality to cancel out those parameters that are constant with respect to ϕ_0:

$$\begin{aligned} p(\phi_0 \mid Y, Z, \omega_1) &\propto p(Y \mid Z, \phi_0, \omega_1) \times \pi(\phi_0) \times \pi(Z \mid \omega_1) \times \pi(\omega_1) \\ &\propto \prod_{i}^{G} p(r_i, s_i \mid Z_i, \phi_0, \omega_1) \times \pi(\phi_0) \times \pi(Z \mid \omega_1) \times \pi(\omega_1) \\ &\propto \prod_{i=1}^{non} \text{Gumbel}(r_i \mid m, \tau) \times \pi(\phi_0) \end{aligned} \tag{8}$$

Similarly, we derive a conditional distribution for the individual essentiality values Z_i, specifying both possible essentiality assignments (i.e., $Z_i = 1$ and $Z_i = 0$):

$$\begin{aligned} &p(Z_i = 1 \mid Y, Z_{\{-i\}}, \phi_0, \omega_1) \\ &\propto p(s_i \mid Z_i = 1) \times p(r_i \mid s_i, Z_i = 1) \times \pi(Z_i = 1 \mid \omega_1) \\ &\propto \Omega(s_i) \times \text{N}(r_i - \lambda_s s_i, \sigma_s^2) \times \omega_1^{Z_i=1}(1 - \omega_1)^{1-Z_i=1} \end{aligned} \tag{9}$$

$$\begin{aligned} &p(Z_i = 0 \mid Y, Z_{\{-i\}}, \phi_0) \\ &\propto p(r_i \mid Z_i = 0, \phi_0) \times p(s_i \mid r_i, Z_i = 0) \times \pi(Z_i = 0 \mid \omega_1) \\ &\propto \text{Gumbel}(r_i \mid m, \tau) \times \text{N}(s_i - \lambda_r r_i) \times \omega_1^{Z_i=0}(1 - \omega_1)^{1-Z_i=0} \end{aligned} \tag{10}$$

2.2 Metropolis-Hastings Sampling

We wish to obtain posterior estimates of essentiality for all genes, integrating over possible values of the unknown variables (e.g. ϕ_0). In order to accomplish this, we generate a Markov-Chain Monte-Carlo sample of values from the conditional densities of interest. By sampling from these conditional densities we can obtain posterior estimates of essentiality, Z_i, without having to know or calculate the probability of non-insertions, ϕ_0, before hand; effectively integrating over this parameter.

Because the conditional distribution of Z_i admits only two possible outcomes (i.e., essential and non-essential), this density can be sampled from a Bernoulli distribution with outcomes proportional to the normalized conditional probability (9, 10):

$$Z_i^{(j)} \sim \text{Bernoulli}(\frac{p_1}{p_1 + p_0})$$

$$\begin{aligned} p_1 &= p(r_i, s_i|Z_{\{-i\}}, \phi_0) \times \omega_1 \\ p_0 &= p(r_i, s_i|Z_{\{-i\}}, \phi_0) \times (1 - \omega_1) \end{aligned}$$

However, the posterior distribution for the parameter ϕ_0 (8) is the product of multiple Gumbel distributions and a Beta distribution, which are not conjugate with each other, and cannot be easily sampled. In order to sample from this posterior density, we utilize a random-walk Metropolis Hastings (MH) algorithm. The MH algorithm is capable of sampling from arbitrary distributions of interest by proposing new candidate values from a Gaussian distribution centered on the last accepted value, $\phi_0^{(j-1)}$, with small variance, v, and accepting or rejecting candidate values probabilistically. Algorithm 1 presents the sampling scheme used to sample the posterior densities of ϕ_0 and Z_i. A MH step is taken to sample ϕ_0, and then we sample Z_i for each gene.

Algorithm: Random-Walk Metropolis-Hastings

Result: MCMC Samples of density $p(Z|Y, \phi_0)$ and $p(\phi_0|Y, Z)$
Assign starting value to ϕ_0, and initialize Z based on proportion of insertions within individual genes (i.e. If $\frac{|TA|_i}{n_i} < 0.1$ then $Z_i = 1$ else $Z_i = 0$);
for *j=1 to desired sample size* **do**
  Draw candidate parameter ϕ_0^c from Normal distribution, $\text{N}(\phi_0^{j-1}, v)$;
  Compute ratio R = $\frac{p(\phi_0^c|Y,Z)}{p(\phi_0^{j-1}|Y,Z)}$;
  Draw u from uniform distribution on [0,1] ;
  if $u < R$ **then**
    Set $\phi_0^{(j)} = \phi_0^c$;
  else
    Set $\phi_0^{(j)} = \phi_0^{j-1}$;
  end
  Let K_z equal the number of genes with $Z_i^j = 1$;
  Let G be the total number of genes;
  Sample $\omega_1^{(j)} \sim \text{Beta}(\alpha_w + K_z, \beta_w + G - K_z)$;
  for $i \leftarrow 1$ **to** G **do**
    $p_1 = p(r_i, s_i|Z_i = 1, Z_{\{-i\}}, \phi_0) \times \omega_1$;
    $p_0 = p(r_i, s_i|z_i = 0, Z_{\{-i\}}, \phi_0) \times (1 - \omega_1)$;
    Sample $Z_i^{(j)} \sim \text{Bernoulli}(\frac{p_1}{p_1+p_0})$;
  end
end

Algorithm 1: Random-Walk Metropolis-Hastings Algorithm for Sampling ϕ_0 and Z

After the samples of parameter value, ϕ_0, and essentiality assignments, Z_i, are obtained, their posterior estimates can be obtained by averaging over the final sample, minus a burn-in stage to ensure the sampling procedure has mixed well.

3 RESULTS

We applied our method to deep-sequencing data from transposon-insertion libraries of the H37Rv strain of *M. tuberculosis*. The full details of the construction of this library is presented in Griffin *et al.* (2011). Briefly, the libraries were prepared by transforming H37Rv using the MycoMarT7 phage, leading to approximately 10^5 independent insertion events. Colony forming units were inoculated into 200 ml of minimal media and 0.1% glycerol, and grown at 37°C. The libraries were sequenced with an Illumina GAII sequencer, and a read length of 36 bp (6-8 million reads per library).

The H37Rv genome has 4,411,654 bp, and contains a total of 3,989 open reading frames (ORFs) (Cole *et al.*, 1998). This equates to an average of 15.9 TA sites per ORF, spaced roughly 61 bp apart on average. Reads from two independent libraries were obtained, which were then summed together to achieve a higher sampling density of the TA sites. Of the 74,605 total TA sites in the genome, 38,984 (53.12%) had reads mapping to them, showing evidence of a transposon insertion at those locations. Of these insertion sites, 32,701 of them occurred within ORFs. We assume that sites with a small number of reads (i.e., 1) could represent spurious reads possibly due to sequencing errors, and therefore those sites were treated as lacking any insertions. However, Figure S9 shows the read counts fit an overdispersed Poisson distribution, suggesting most are legitimate insertions. Sites with just 1 insertion are discarded anyway to be safe, requiring insertions to be confirmed by at least 2 reads. This might lower the effective density of the dataset, however this does not affect the method. Of the 3989 ORFs in H37Rv, 41 do not contain any TA sites. An additional set of 237 genes were deemed too short due to the fact that they do not contain enough TA sites (i.e, $n_i < 3$) or the span of nucleotides was too short (i.e. $s_i < 150bp$). Therefore a total of 278 genes are reported as 'No-Data' because our analysis is not appropriate for these genes.

The sampling procedure was run for 50,000 iterations, providing estimates of essentiality for all viable genes, as well as estimates of the parameter ϕ_0. To ensure that the algorithm mixed well and the samples obtained were uncorrelated, the first 1,000 samples were treated as a "burn-in" period and discarded; only keeping every 20th sample after there. Section 2 of the Supplementary Material contains an analysis of the convergence of the Metropolis-Hastings procedure used. The value for ϕ_0 (non-insertion frequency in non-essential genes) was estimated to be 0.290 ± 0.004 (std. dev.). Performance on a lower-density dataset, also H37Rv grown on glycerol, is described in Section 3 of the Supplementary Material. This lower-density library contains fewer transposon insertions in coding regions (i.e., 23,399 (36.3%) compared to 31,715 (50.4%) in the library described above), and has longer runs of non-insertions among the genes ($\phi_0 = 0.592$). Rather than predicting more essential genes, our analysis is more conservative in its predictions as it is less confident of the essentiality of the genes given the sparsity of the insertions.

3.1 Essentiality Results

After obtaining the sample from the MH procedure, the posterior probability of essentiality for all genes is estimated by averaging over the sample of essentiality values, $\bar{Z}_i$. To set significance thresholds while correcting for multiple comparisons (i.e. to control false discovery rate, FDR), we utilize a method described by Muller *et al.* (2006) which emulates the Benjamini-Hochberg procedure for Bayesian posterior probabilities. Limiting the FDR at 0.05, genes with $\bar{Z}_i < 0.0371$ are classified as non-essential, and genes with $\bar{Z}_i > 0.9902$ are classified as essential. Table S1 of the Supplementary Material contains our predictions for all 3989 ORFs in H37Rv. In total, 667 genes are categorized as essential and 2693 are non-essential by this criterion. These include genes experimentally validated to be essential for growth in vitro (i.e. *prrA* and *prrB* (Haydel *et al.*, 2012), *phoP* (Goyal *et al.*, 2011), and *mshA* and *mshC* (Buchmeier and Fahey, 2006)) and genes known to be non-essential (i.e. *rpfA* (Kana *et al.*, 2008), *glnD* (Read *et al.*, 2007), *echA5* and *fadB3* (Williams *et al.*, 2011)) The remaining genes represent those for which the method is unable to reach an essentiality assignment with confidence. Figure 2 shows a cumulative plot of the $\bar{Z}_i$ values for all the genes, with the blue lines representing the thresholds of essentiality and non-essentiality. To assess the sensitivity of this result to the fixed parameters in the likelihood function for essential genes, we obtained results for different values of of δ and κ parameters of the sigmoid function. Figure S2 of the Supplementary Material shows a cumulative plot of $\bar{Z}_i$ values for different combinations of these parameters. The κ parameter has little effect on the final result. On the other hand, a two-fold increase and decrease of the δ parameter significantly changes the slope of the graph as well as then number of non-essential genes estimated. This is consistent with the fact that the δ parameter represents the expected span of nucleotides for essential domains. This has been empirically determined to be approximately 300 nucleotides.

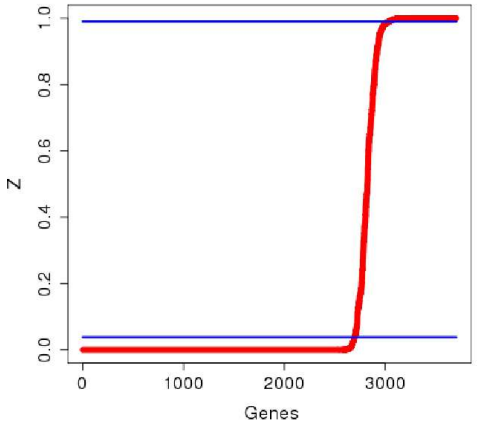


Fig. 2. Cumulative plot of posterior probabilities $\bar{Z}_i$. The average Z_i value for each gene was plotted in ascending order, for each of the different combinations of parameters investigated. The blue lines represent the final thresholds for essentiality: $\bar{Z}_i > 0.9902$ and $\bar{Z}_i < 0.0371$

Table 1 reports statistics for the different categories of genes. On average essential genes contained significantly longer maximum runs of non-insertion (17.57) than non-essential genes and these runs spanned a larger number of nucleotides (1039.81 bp), which is consistent with our expectations for essentiality. Non-essential genes contained a larger number of insertions on average (15.69). Although essential genes contained only a small number of insertions (1.68) this number was greater than zero, indicating that the method is capable of detecting essential genes with a small

number of insertions, provided they contain a long enough run of non-insertions suggestive of an essential region.

Table 1. Statistics for Essentials, Non-Essentials and Uncertain Genes. Non-Essential genes are those with $Z_i < 0.0371$, Essential genes are those with $Z_i > 0.9902$. Average span is in nucleotides.

	Total	Average			
	Genes	TA Sites	Insertions	Max Run	Span
Essentials	667	21.32	1.68	17.57	1039.8
Uncertain	342	16.61	6.45	5.75	410.4
Non-Essentials	2693	15.69	10.78	2.05	54.5

3.2 Concordance with Previous Results

The essentiality of the entire *M. tuberculosis* H37Rv genome has been characterized previously using transposon-site hybridization (Sassetti *et al.*, 2003; Sassetti and Rubin, 2003). We compare our essentiality inferences to previous results to verify that our method achieves results that are consistent with expectations of the essentiality in *M. tuberculosis*. Sassetti *et al.* utilized Transposon Site Hybridization (TraSH) to characterize the genes necessary for optimal growth *in vitro*, for a library of transposon mutants grown on 0.02% glucose and rich-media (7H10). While our method analyzes deep sequencing of transposon libraries, TraSH utilizes hybridization of gene-specific probes to quantify the level of fluorescence being emitted by hybridization probes to determine which genes are being interrupted in the library of mutants. Table 2 contains a comparison between the two methods.

Table 2. Comparison of Essentiality Predictions with TraSH analysis. The results obtained by Sassetti. et. al are compare with those obtained with our Bayesian method for all 3989 genes in *M. tuberculosis*

		Bayesian Method				
		Essential	Uncertain	Non-Essential	No-Data	Total
Sassetti-03	**Essentials**	429	75	81	29	614
	Growth-Defect	9	4	28	1	42
	Non-Essential	94	151	2131	144	2520
	No-Data	135	112	453	113	813
	Total	667	342	2693	287	3989

Sassetti *et al.* also included an additional category of genes representing those whose interruption causes growth-defects (i.e. slower growth); our method does not make this distinction. Excluding these, the two methods show agreement in 69.9% of essentials, and 84.6% of non-essentials for a total of 81.7% across both categories. There were only 81 genes predicted to be essential by TraSH but not by our method, and 94 genes predicted to be non-essential by TraSH but found to be essential by our method.

Some of these differences could be due to the different growth conditions of the libraries. For example, because our library was grown on glycerol we find genes necessary for glycerol metabolism as essential, such as GlpK (glycerol kinase). Other differences may be due to incomplete sequence coverage (e.g. gaps in PE_PGRS genes, which are highly GC-rich and hard to sequence). Two out of 62 PE_PGRS genes in the H37Rv genome were classified as essential by our model because of large regions without insertions, though genes in this family are generally believed to be non-essential (Banu *et al.*, 2002). Over-representation of PE_PGRS gene among essentials was also noted in other transposon library analyses using sequencing (Lamichhane *et al.*, 2003).

One notable difference is that Sassetti *et al.* found *glcB* to be non-essential, however the insertion pattern shown in Figure 3 clearly indicates that this gene was unable to tolerate insertions in the libraries of mutants analyzed. GlcB encodes for malate synthase in *M. tuberculosis*, which was originally thought to be necessary only for growth on fatty-acids as part of a glyoxylate shunt (McKinney *et al.*, 2000), but has recently been shown to be essential on other carbon sources like dextrose by chemical inhibition (Krieger et al., accepted). A complete absence of transposon insertions in Rv1837c was also observed in the DeADMAn studies (Lamichhane *et al.*, 2003). Our data suggests that GlcB is also essential for growth on glycerol (in liquid culture with minimal media), showing a significant run of non-insertions (25 out of 27 - spanning 2078 nucleotides, p($Z_i = 1$)=1.0). It should be be noted that in the original TraSH data, GlcB had a hybridization ratio of 0.41, which was near the threshold for essentiality (< 0.20).

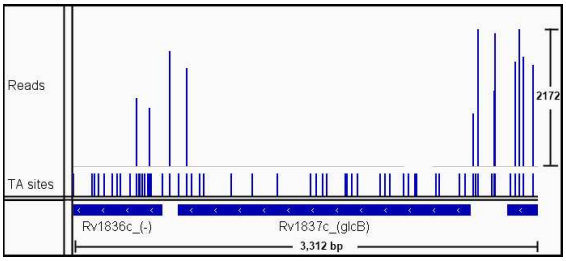


Fig. 3. Insertion Pattern for Rv1837c. Figures created using IGV - distributed by the Broad Inst. http://www.broadinstitute.org/igv/

3.3 Comparison to Other Statistical Models

In contrast to other models where the order of insertions does not matter, the Gumbel distribution is capable of identifying regions lacking insertions within genes that are significantly longer than expected, despite the presence of insertions elsewhere in the gene. Models of essentiality that focus solely on the proportion of insertions or the number of reads within genes, may miss these essential regions if enough insertions are observed elsewhere within the gene.

To evaluate this important distinction, we compare our method to the method proposed by Blades and Broman (2002). This method does not take the order of insertions into consideration but instead is based on a multinomial likelihood function that characterizes the number of mutants with insertions unique to a gene, as well the number of mutants with insertions that occur in regions that overlap with adjacent genes (which adds uncertainty as to which gene was disrupted). Using this multinomial likelihood, Blades and Broman implement a Gibbs Sampling procedure that estimates posterior probabilities of essentiality for all the genes.

We obtained the R package "*negenes*" which contains the implementation of this method maintained by Karl W. Broman (http://www.biostat.wisc.edu/ kbroman/software/). The Gibbs sampler was run on the same H37Rv glycerol dataset analyzed above, for a total of 50,000 iterations. The first 1,000 samples were ignored as part of the burn-in period, and only every 22th sample was kept to remove any auto-correlation in the sampling process. Following Lamichhane *et al.* (2003), we used the number of insertions within the N-terminal 80% of the ORF, as representative of the number of viable mutants with insertions in genes.

After obtaining the probability of essentiality from the Gibbs sampler, we set thresholds for essentiality by correcting for multiple comparisons and controlling the false discovery rate as we did for the Gumbel method (implementing a procedure analogous to the one proposed by Benjamini and Hochberg). We use these thresholds to classify the genes as essential, non-essential or uncertain.

The Blades and Broman method predicts a total of 244 essential genes and 3195 non-essential genes. As the small number of essential genes could be due to the selected threshold, a less conservative threshold for essentiality (posterior probability > 0.95) results in 458 genes being predicted as essential, still well below the 614 essential genes characterized by Sassetti *et al.* (2003). A full breakdown of the results is found in table S5 in the Supplementary Material.

The lower number of essentials predicted by the Blades and Broman method is due to the fact that the presence of even a few insertions in a gene (i.e. ≥ 1) is enough to make it seem non-essential under this multinomial model. For example, GyrB (β subunit of DNA gyrase) a known essential gene and target of fluoroquinolones, is found to be non-essential by the Blades and Broman method as it is observed to have insertions at the N-terminus. In contrast our method finds GyrB to be essential as it contains a significant stretch of TA sites lacking any insertions (40 consecutive TA sites lacking insertions out of a total of 43 TA sites in the gene) which is what we would expect from an essential gene. Although this could be potentially overcome by ignoring insertions at the very ends of the N- and C- termini (for example, considering only 5-80% of the coding region as recommended by Gawronski *et al.* (2009)), a strength of our model is that it does not need to discard these regions a priori.

The dataset was also analyzed with the ESSENTIALS software, which utilizes a negative binomial distribution to analyze read counts within each gene (Zomer *et al.*, 2012). ESSENTIALS only predicts 434 genes in the H37Rv genome to be essential (using a -8.13-fold-change cutoff determined by the software), and 3363 genes to be non-essential. Thus, similar to the Blades and Broman method, the ESSENTIALS software also under-predicts the number of essential genes compared to what is expected for this organism (i.e. $\sim$600, based on prior TraSH studies). However, 93.5% of the genes predicted to be essential by the ESSENTIALS software (406/434) overlap with the genes predicted to be essential by our Bayesian method, effectively representing a subset (60% of our 667). 311 of the 434 genes were correctly labeled as essential according to the original TraSH experiments, capturing 50% of the 614 essential genes previously characterized. The primary reason that other genes that are believed to be essential are not identified as such by the ESSENTIALS software appears to be that many these genes often contain some insertions at a few sites in the ORF (such as at the N- and C- terminii, or in a non-essential domain), and thus their read counts are higher than expected for an essential gene (according to their model).

3.4 Essential Domains

One of the advantages of our method for analyzing this high-resolution transposon insertion data is that it can reveal essential regions or domains within proteins. The regions devoid of insertions detected by our method often correspond to well-defined protein folding domains. To date, the X-ray crystallographic structures of only 8.5% of proteins in H37Rv have been determined (Ehebauer and Wilmanns, 2011). Thus, in order to test the model's ability to detect essential domains, we compare these regions to Pfam predictions of protein domains within H37Rv genes. Pfam is able to make predictions of domains based on amino-acid sequence homology using Hidden Markov Models to represent protein families (Finn *et al.*, 2010).

Using Pfam, we obtained predictions of a total 5,091 protein domains, 1,126 of which were in genes predicted to be essential by our method. In order to determine whether the significant runs of non-insertions observed coincide with protein domain boundaries, we calculated a ratio of overlap between the nucleotides spanned by a maximum run of non-insertions and the domain boundaries predicted by Pfam (Table 3). For a Pfam domain spanning coordinates $i..j$, if the closest matching region lacking insertions is $k..l$, then the overlap score is $min(|i-k|, |j-l|)/|i-j|$. Of the 1,126 domains found within essential genes, 976 (86.68%) of them overlapped significantly (i.e. score > 0.80) with the domain boundaries, suggesting that the majority of Pfam domains are contained within the runs of non-insertions observed. Another set of 104 domains (9.24%) had no significant overlap (i.e. score < 0.20), potentially representing those domains that are within non-essential regions of essential genes. The remaining 46 (4.08%) domains, represent those with an intermediate overlap (i.e. score between $0.20-0.80$), representing a small set of genes for which the Pfam boundary prediction may be inconsistent.

Table 3. Statistics for Pfam Domain Predictions.

	Total	Average		
	Domains	**Length**	**TA sites**	**Overlap**
Non-Essentials	3240	464.45	7.47	0.28
Uncertain	512	457.90	6.30	0.47
Essentials	1126	518.78	8.31	0.89

In some cases, the pattern of transposon insertions is capable of identifying individual domains as essential through a sequence of TA sites lacking insertions that closely matches the boundaries of the predicted Pfam domain. To identify such cases, we matched the Pfam domain predictions to the closest run of non-insertions, and calculated a consistency score that reflects the consistency between the two regions. This consistency score was based on comparing the distance (in nucleotides) between the boundaries of the domain prediction $i..j$ and the boundaries of the run of non-insertions $k..l$. We restricted attention to genes for which the run of

non-insertions corresponding to the domain is statistically significant (p<0.05 using a cumulative Gumbel distribution), and whose distance between boundaries (upstream or downstream) is less than 50 bp (i.e. $|i-k|+|j-l|<50$). We identified 95 known domains which were mostly devoid of transposon insertions internally, but for which insertions were observed at TA sites right near the boundaries (shown in Table S2 in the Supplementary Material). Many of these constitute essential single-domain proteins, although several occur in larger multi-domain proteins with both essential and non-essential regions. To identify genes that contain both essential and non-essential domains, we selected a subset of genes that are labeled as essential by our Bayesian analysis, but for which there is still a relatively large (i.e. #TAs >4) area remaining containing an insertion frequency that is not significantly essential (i.e. p-value >0.05) according to the cumulative Binomial distribution. This gave a set of 36 genes (presented in Table S3 of the Supplementary Material) that represent interesting cases where there is a combination of both essential and non-essential regions.

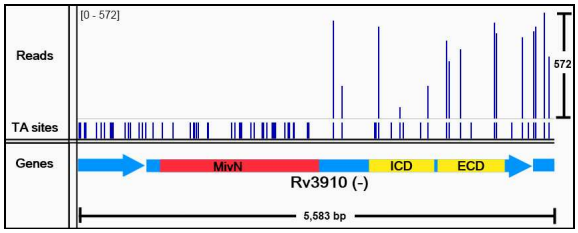


Fig. 4. Insertion Pattern for Rv3910. The essential MviN domain is shown in red, while the non-essential extracellular and intracellular domains are shown in yellow.

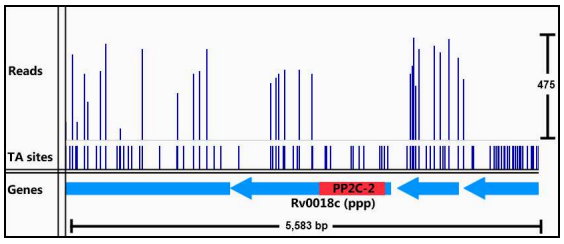


Fig. 5. Insertion Pattern for Rv0018c. The essential catalytic domain is shown in red.

Among the genes that our method identifies as having both essential and non-essential domains are Rv3910 and Rv0018c. These two genes have been shown to be essential for growth in mycobacteria, and are involved in regulating cell wall (peptidoglycan) synthesis. Rv3910 encodes for two C-terminal protein domains (an intracellular pseudokinase and an extracellular sugar-binding domain) and a N-terminal, MviN-like, domain which is required for the late-stages of peptidoglycan biosynthesis (Figure 4). MviN proteins have been proposed to be involved in the export of the lipid-II precursor and this *Mtb* ortholog is regulated by a phosphorylation-dependent interaction with FhaA (Rv0020c) (Gee *et al.*, 2012). Insertions in Rv3910 are found only in the C-terminal domains, but not the N-terminal membrane domain, implying only the latter domain is necessary for growth ($p(Z_i=1)=1.00$)), based on a run of 34 consecutive non-insertions within its essential domain, spanning 1439 nucleotides, as significant. This has been confirmed experimentally by Gee *et al.* (2012). Rv0018c (PstP: serine/threonine phosphatase) contains an essential catalytic domain within its N-terminus (Figure 5), and has been shown to dephosphorylate Rv0020 (FhaA) counteracting phosphorylation by PknB (Pullen *et al.*, 2004). It contains a run of 12 consecutive non-insertions in the N-terminal catalytic domain, spanning 695 nucleotides, which our method is identifies as significant ($p(Z_i=1)=0.999$)).

4 DISCUSSION

The availability of next generation sequencing data for analyzing transposon mutagenesis libraries necessitates a new method capable of analyzing the high-resolution data and determining essentiality. We developed a Bayesian statistical model that can be used to analyze this data and make rigorous predictions about the essentiality of individual genes, as well as identify stretches of non-insertions indicative of essential domains. Using this method we have analyzed sequence data from a library of mutants of *M. tuberculosis* bacteria, and found high concordance with previous results.

The key insight in our method is the use of the Gumbel distribution to assess the statistical significance of runs of non-insertion that are significantly longer than expected, and therefore indicative of essential regions. By highlighting these essential regions, our method can help identify essential domains within genes which might otherwise be missed by characterizing the proportion of insertions alone. An approach based on analyzing the size of insertion gaps was suggested by Christen *et al.* (2011), however they used a simpler exponential model for assessing statistical significance. Using our method we found multiple genes with essential domains that match Pfam predictions of domains, and whose essentiality is supported in the literature (e.g. Rv0018c, Rv3910). Moreover, because our method depends on consecutive sub-sequences of TA sites lacking insertions, and not on the simple presence or absence of insertions within a gene, our method is not sensitive to insertions at the N- or C- terminus of a gene, which essential genes have been shown to occasionally tolerate (Akerley *et al.*, 1998; Smith *et al.*, 1996; Christen *et al.*, 2011).

Although previous analyses have used deep-sequencing data to determine essentiality, these have relied on *ad hoc* criteria to ignore insertions at the N- or C- terminus, or made assumptions about parameters to quantify the confidence of their essentiality predictions. However, by using a Bayesian statistical framework, our method can simultaneously estimate model parameters and posterior probabilities of essentiality without requiring *a priori* estimates of these unknown variables.

While our method can successfully determine regions in the genome that contain unusually long gaps lacking any insertions, it does not take the number of reads observed at each site into account (i.e. read counts). Our binary interpretation of the insertion data is based on mere presence or absence of reads, and ignores any potential information that the number of reads mapping to a particular site might bear. An alternative approach based on quantifying read counts within genomic regions might yield biologically-relevant information on essentiality. For example, in another paper (Zhang *et al.*, 2012), a model was developed which calculates significance

scores for sums of read counts at TA sites using a non-parametric test. This approach yields qualitatively similar results to our model in terms of which genes are classified as essential and non-essential. One advantage of an approach based on read counts is that it could potentially detect genes whose disruption leads to growth defects, in that slower-growing mutants might produce fewer (but non-zero) read in a disrupted gene compared to the expected value. This was exploited to infer genes that play a role in cholesterol catabolism by Griffin *et al.* (2011). On the other hand, a single over-amplified TA site in a region (with an excess of reads due to PCR bias) could lead to misinterpretation, whereas our model would be less sensitive to this kind of noise.

The method we have presented can be used to assess essentiality of genes, as well as inter-genic (e.g. regulatory) regions and assign statistical confidence scores, in any organism, provided a transposon mutant library can be constructed and sequenced. Although mutant libraries analyzed were constructed using the *Himar1* transposon, this method could be used to analyze libraries constructed using the Tn5 transposon, where every nucleotide is a possible insertion site (Langridge *et al.*, 2009). By sampling the probability of non-insertion from its posterior distribution, we can calculate estimates of essentiality for a diversity of mutagenesis experiments.

ACKNOWLEDGEMENTS

Funding: This work was support in part by funding from the Robert A. Welch Foundation, grant number A-0015 (JCS), and NIH grant U01GM094568 (JCS, TRI, EJR).

REFERENCES

Akerley, B. J., Rubin, E. J., Camilli, A., Lampe, D. J., Robertson, H. M., and Mekalanos, J. J. (1998). Systematic identification of essential genes by in vitro mariner mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 8927–8932.

Banu, S., Honore, N., Saint-Joanis, B., Philpott, D., Prevost, M. C., and Cole, S. T. (2002). Are the PE-PGRS proteins of Mycobacterium tuberculosis variable surface antigens? *Mol. Microbiol.*, **44**, 9–19.

Blades, N. J. and Broman, K. W. (2002). Estimating the number of essential genes in a genome by random transposon mutagenesis. Technical Report MSU-CSE-00-2, Dept. of Biostatistics Working Papers, Johns Hopkins University.

Buchmeier, N. and Fahey, R. C. (2006). The mshA gene encoding the glycosyltransferase of mycothiol biosynthesis is essential in Mycobacterium tuberculosis Erdman. *FEMS Microbiol. Lett.*, **264**(1), 74–79.

Christen, B., Abeliuk, E., Collier, J. M., Kalogeraki, V. S., Passarelli, B., Coller, J. A., Fero, M. J., McAdams, H. H., and Shapiro, L. (2011). The essential genome of a bacterium. *Mol. Syst. Biol.*, **7**, 528.

Cole, S. T., Brosch, R., and Parkhill, J. (1998). Deciphering the biology of mycobacterium tuberculosis from the complete genome sequence. *Nature*, **393**(6685), 537–544.

Ehebauer, M. T. and Wilmanns, M. (2011). The progress made in determining the Mycobacterium tuberculosis structural proteome. *Proteomics*, **11**, 3128–3133.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R., and Bateman, A. (2010). The Pfam protein families database. *Nucleic Acids Res.*, **38**, D211–222.

Gawronski, J. D., Wong, S. M. S., Giannoukos, G., Ward, D. V., and Akerley, B. J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for haemophilus genes required in the lung. *PNAS*, **106**(38), 16422–16427.

Gee, C. L., Papavinasasundaram, K. G., Blair, S. R., Baer, C. E., Falick, A., King, D., Griffin, J. E., Venghatakrishnan, H., Zukauskas, A., Wei, J., Dhiman, R., Crick, D., Rubin, E., Sassetti, C. M., and Alber, T. (2012). A phosphorylated pseudokinase complex controls cell wall synthesis in mycobacteria. *Sci Signal*, **5**, ra7.

Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balazsi, G., and et al. (2003). Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J. Bacteriol.*, **185**(19), 5673–5684.

Goyal, R., Das, A. K., Singh, R., Singh, P. K., Korpole, S., and Sarkar, D. (2011). Phosphorylation of PhoP protein plays direct regulatory role in lipid biosynthesis of Mycobacterium tuberculosis. *J. Biol. Chem.*, **286**(52), 45197–45208.

Griffin, J. E., Gawronski, J. D., DeJesus, M. A., Ioerger, T. R., Akerley, B. J., and Sassetti, C. M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog*, **7**(9), e1002251.

Hasan, S., Daugelat, S., Rao, P. S., and Schreiber, M. (2006). Prioritizing genomic drug targets in pathogens: application to Mycobacterium tuberculosis. *PLoS Comput. Biol.*, **2**(6), e61.

Haydel, S. E., Malhotra, V., Cornelison, G. L., and Clark-Curtiss, J. E. (2012). The prrAB two-component system is essential for Mycobacterium tuberculosis viability and is induced under nitrogen-limiting conditions. *J. Bacteriol.*, **194**(2), 354–361.

Hayes, F. (2003). Transposon-based strategies for microbial functional genomics and proteomics. *Annu. Rev. Genet.*, **37**, 3–29.

Kana, B. D., Gordhan, B. G., Downing, K. J., Sung, N., Vostroktunova, G., Machowski, E. E., Tsenova, L., Young, M., Kaprelyants, A., Kaplan, G., and Mizrahi, V. (2008). The resuscitation-promoting factors of Mycobacterium tuberculosis are required for virulence and resuscitation from dormancy but are collectively dispensable for growth in vitro. *Mol. Microbiol.*, **67**(3), 672–684.

Krieger, I., Freundlich, J., Gawandi, V., Roberts, J., Gawandi, V., Sun, Q., Owen, J., Fraile, M., Huss, S., Lavandera, J.-L., Ioerger, T., and Sacchettini, J. (2012). Structure-guided discovery of phenyl diketo-acids as potent inhibitors of M. tuberculosis malate synthase. *Chemistry & Biology*.

Lamichhane, G., Zignol, M., Blades, N. J., Geiman, D. E., Dougherty, A., Grosset, J., Broman, K. W., and Bishai, W. R. (2003). A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: Application to mycobacterium tuberculosis. *PNAS*, **100**(12), 7213–7218.

Lamichhane, G., Tyagi, S., and Bishai, W. R. (2005). Designer arrays for defined mutant analysis to detect genes essential for survival of Mycobacterium tuberculosis in mouse lungs. *Infect. Immun.*, **73**(4), 2533–2540.

Lampe, D. J., Churchill, M. E., and Robertson, H. M. (1996). A purified mariner transposase is sufficient to mediate transposition in vitro. *The European Molecular Biology Organization Journal*, **15**(19), 5470–5479.

Langridge, G. C., Phan, M.-D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E., Dougan, G.,

and et al. (2009). Simultaneous assay of every salmonella typhi gene using one million transposon mutants. *Genome Research*, **19**(12), 2308–2316.

McKinney, J. D., Honer zu Bentrup, K., Munoz-Elias, E. J., Miczak, A., Chen, B., Chan, W. T., Swenson, D., Sacchettini, J. C., Jacobs, W. R., and Russell, D. G. (2000). Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature*, **406**, 735–738.

Muller, P., Parmigiani, G., and Rice, K. (2006). Fdr and bayesian multiple comparisons rules. In *Proceedings of the ISBA 8th World Meeting on Bayesian Statistics*, Benidorm, Spain.

Pullen, K. E., Ng, H. L., Sung, P. Y., Good, M. C., Smith, S. M., and Alber, T. (2004). An alternate conformation and a third metal in PstP/Ppp, the M. tuberculosis PP2C-Family Ser/Thr protein phosphatase. *Structure*, **12**, 1947–1954.

Read, R., Pashley, C. A., Smith, D., and Parish, T. (2007). The role of GlnD in ammonia assimilation in Mycobacterium tuberculosis. *Tuberculosis (Edinb)*, **87**(4), 384–390.

Rubin, E. J., Akerley, B. J., Novik, V. N., Lampe, D. J., Husson, R. N., and Mekalanos, J. J. (1999). In vivo transposition of mariner-based elements in enteric bacteria and mycobacteria. *PNAS*, **96**(4), 1645–1650.

Sassetti, C. M. and Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *PNAS*, **100**(22), 12989–12994.

Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology*, **48**(1), 77–84.

Schilling, M. F. (1990). The longest run of heads. *College of Mathematics Journal*, **21**, 196–207.

Smith, V., Chou, K. N., Lashkari, D., Botstein, D., and Brown, P. O. (1996). Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science*, **274**, 2069–2074.

Wheelan, S. J., Marchler-Bauer, A., and Bryant, S. H. (2000). Domain size distributions can predict domain boundaries. *Bioinformatics*, **16**(7), 613–618.

Williams, K. J., Boshoff, H. I., Krishnan, N., Gonzales, J., Schnappinger, D., and Robertson, B. D. (2011). The Mycobacterium tuberculosis -oxidation genes echA5 and fadB3 are dispensable for growth in vitro and in vivo. *Tuberculosis (Edinb)*, **91**(6), 549–555.

Zhang, Y. J., Ioerger, T. R., Huttenhower, C., Long, J. E., Sassetti, C. M., Sacchettini, J. C., and Rubin, E. J. (2012). Global Assessment of Genomic Regions Required for Growth in Mycobacterium tuberculosis. *PLoS Pathog.*, **8**(9), e1002946.

Zomer, A., Burghout, P., Bootsma, H. J., Hermans, P. W., and van Hijum, S. A. (2012). ESSENTIALS: software for rapid analysis of high throughput transposon insertion sequencing data. *PLoS ONE*, **7**(8), e43012.

Chapter 3.

Mycobacterial Determinants for Survival of the CD4 T Cell Response

Section 3.1: Overview and Attributions

Overview. In this chapter we determined the genetic requirements for mycobacterial survival during infection, as well as the specific determinants of surviving the CD4 T cell response. We describe 576 genes required for growth *in vivo*, determined by comparing the insertion profiles of inoculum libraries to the surviving pools. CD4 T cells are crucial in anti-mycobacterial immunity, but are unable to completely clear the pathogen. Thus, we hypothesized that many of these genes required during infection would be responsible for helping Mtb survive CD4-mediated immunity. We screened for these CD4 survival determinants and found 58 genes wherein significantly more insertions were found from the surviving pools in MHC Class II knockout mice (which lack CD4 T cells) compared to wild type mice. Through pathway enrichment and further gene requirement profiling, our results suggest that bacterial tryptophan biosynthesis is specifically required to combat CD4 mediated immunity. We conclude that CD4 T cells induce tryptophan starvation, and Mtb survival of the CD4 response is dependent on its ability to respond to that starvation through tryptophan biosynthesis.

Attributions. The following Chapter is an expanded version a manuscript, which also includes most of Chapter 4. I wrote the manuscript and we aim to submit in the next month. I made the mycobacterial transposon library with help from Ed Long, and performed the mouse infections and bacterial harvests. I made the DNA transposon junction libraries, and Thomas Ioerger and colleagues

sequenced them. The statistical analysis was a joint effort with many collaborators, including Thomas Ioerger, Curtis Huttenhower, Ed Long, Chris Sassetti, Chris Ford, and Sarah Fortune. This was an especially dedicated and talented community, and I am grateful to have been a part of it. It is also an expanding community. Statistical modeling of the data can always be improved and continues to be a work in progress, and Jeremy Rock, Justin Pritchard and Michael DeJesus are spearheading these efforts.

Section 3.2: Genetic determinants of Adaptive Immune Evasion by *Mycobacterium tuberculosis*

Abstract

Tuberculosis remains a key global health threat, claiming 2 million lives each year. A staggering 2 billion people are estimated to be infected with *Mycobacterium tuberculosis* (Mtb), serving as an abundant reservoir for potential disease. The endurance of Mtb as a human pathogen owes in large part to its ability to adapt to host adaptive immunity. Upon infection of the host lung, bacteria divide exponentially until the peak of the CD4 response. The importance of CD4 immunity is further supported by the rapid death of infected mice lacking CD4 cells as by the increased susceptibility to Mtb of HIV-positive patients. While CD4 cells are crucial for surviving Mtb infection, they are unable to clear the bacteria. Here, we seek to identify the bacterial genetic determinants of combating CD4-mediated immunity. We infected both wildtype and MHC Class II KO mice with high-density Mtb mutant libraries. In comparing the surviving mutants from both groups, we identified Mtb genes that are required for bacterial survival in wildtype mice but not in mice lacking CD4 cells. Among these are genes that encode for gluconeogenesis and tryptophan synthesis, suggesting that CD4 cells mediate key differences in metabolite availability to the bacterium.

Introduction

Mycobacterium tuberculosis (Mtb), the etiologic agent of tuberculosis (TB), is one of the world's most successful bacterial pathogens. After a centuries-long decline in TB, the last few decades has seen a resurgence in TB, with an estimated 2 billion people infected and about 1.7 million deaths per year[1]. The success of Mtb as a pathogen lies in its specific adaptation to the human host and to the many arms of anti-bacterial immunity[2]. Its ability to survive host defenses is directly responsible for the large reservoir of infected people, and its ability to subvert bactericidal mechanisms allows it to replicate *in vivo* and cause disease[3]. Elucidating these mechanisms will help us understand the complex host-pathogen interface, and targeting these mechanisms will help us develop therapeutics that help the immune system kill Mtb.

Of the components of host immunity to Mtb, CD4 T cells have emerged as one of the most biologically and epidemiologically important. While neither humans nor mice, the most commonly used infectious model for Mtb, can completely clear Mtb during infection, both are able to limit bacterial growth, and in the case of immunocompetent humans, prevent disease[2, 4]. This response is dependent on CD4 T cells. The CD4-deficient MHC Class II knockout mice, and other mice that lack CD4 T cells, cannot stop mycobacterial growth and succumb to disease and death about a year ahead of their wildtype counterparts[5-9]. In human disease, progressive CD4 T cell loss due to HIV infection also increases the risk for TB disease and death[10-13].

However, crucial as they are for normal TB immunity, CD4 T cells ultimately fail to kill all bacilli. Survivors remain latent, with the potential to cause disease in the future[2, 4]. The nature of the environment imposed by CD4 T cells, enough to limit growth but not kill Mtb, is not well characterized. Reports have shown that the Th1 subset is especially effective in limiting Mtb growth, and cytokines such as IFN-γ and TNF-α are needed in certain models, though not all, of CD4 T cell mediated defenses[8, 14-19]. But the exact nature of CD4-mediated stress, the repertoire anti-pathogen effectors that the bacillus must survive, is poorly understood.

In this study, we profile the mycobacterial requirements for surviving the CD4 response. 58 genes were shown to be required for mycobacterial survival of the CD4 response. Two pathways were enriched in these genes, including gluconeogenesis and tryptophan biosynthesis, suggesting that CD4 cells altered the *in vivo* metabolite availability for Mtb. In addition to describing the nature of CD4-mediated stress, the results of this Chapter also identify potential drug targets whose inhibition could result in restoration of susceptibility to CD4 immunity.

Results

Genes required for survival during infection

In order to define the set of genes required for surviving the CD4 T cell response, we infected both wildtype and MHC Class II knockout ($MHCII^{-/-}$) mice with a library of Mtb transposon mutants[20]. We injected 10^6 bacteria intravenously and plated surviving bacteria from infected spleens at 10 days and 45 days after infection (Fig 3.1A). To identify mutant Mtb in the surviving pools, we deep sequenced transposon junctions to map the insertion site of each mutant. For each time point, we made two comparisons. First, we compared the surviving pool of mutants from wildtype mice (wt output library) to the inoculating pool (input library), defining genes with a statistically validated decrease in output library as required for growth during infection. Second, we compared the output library from wt mice to the output library from $MHCII^{-/-}$ mice. We reasoned that genes required for surviving the CD4 T cell response would be required for growth in wt mice, while mutants with disruptions in those genes would be rescued in $MHCII^{-/-}$ mice, which lack CD4 T cells (Fig 3.1B).

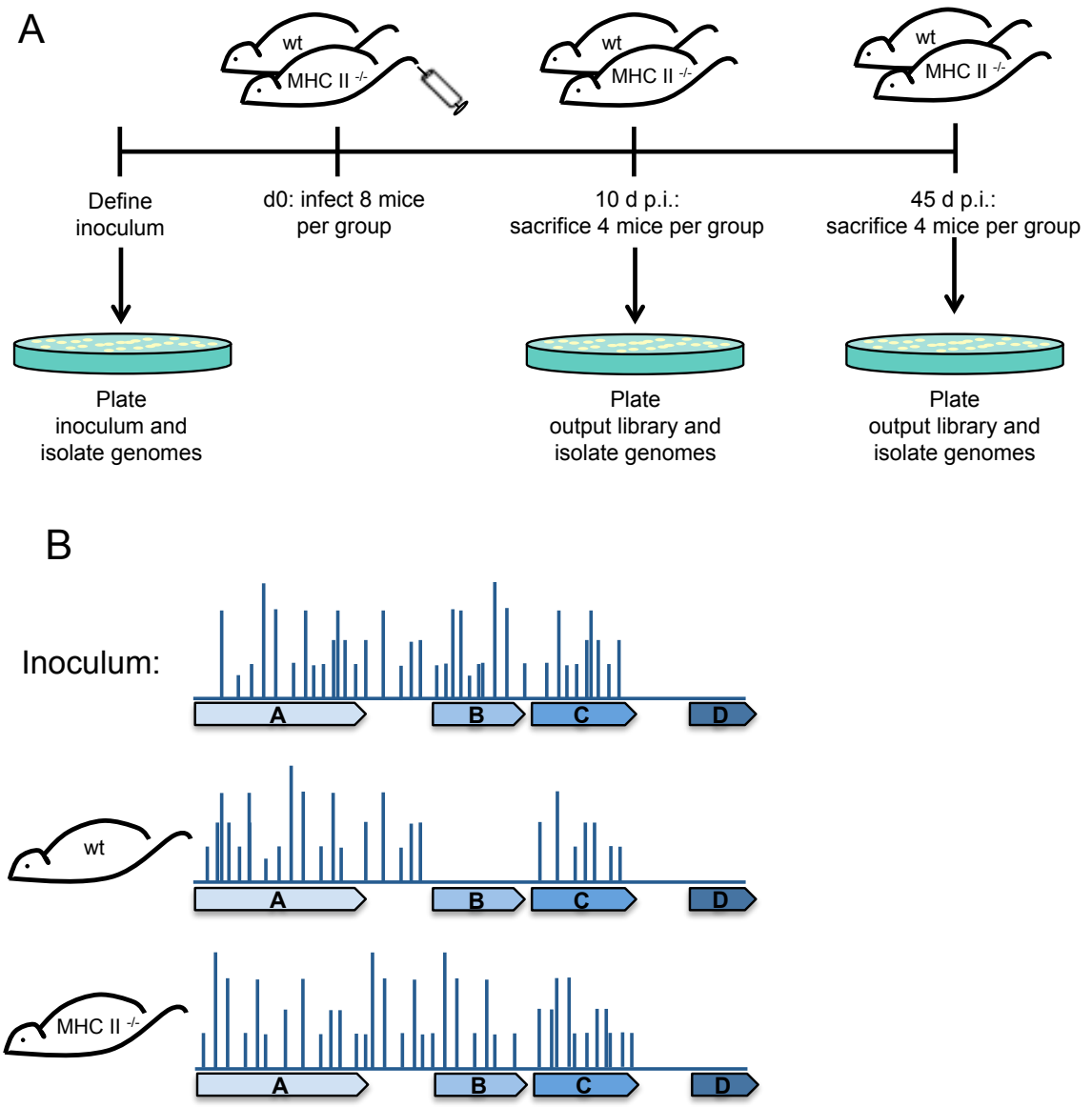


Figure 3.1. A.) Wildtype and MHC Class II knockout mice were infected via intravenous injection with 10^6 CFU of the Mtb transposon library. The inoculum was plated on the day of infection. On days 10 and 45 post-infection, 4 mice in each group were sacrificed and surviving mutants from the spleens were plated. Transposon junctions were prepped and deep sequenced from all plated libraries. B.) We searched for differentially required genes by comparing insertion counts. Genes A and C are examples of non-essential genes that are not required for growth in vivo. Gene B is an example of a gene required for growth in vivo but not required for growth in MHCII$^{-/-}$ mice. Gene D is an essential gene.

We found 576 genes that were required for growth during infection (Fig 3.2). These genes had a statistically significant (FDR q-value < 0.01) 10-fold or more decrease in insertion counts across the gene. On d10, 405 genes were required for growth *in vivo*, and on d45, 317 genes were required. A total of 146 genes were required at both time points (Fig 3.2A and B, Table 3.1). Genes required late but not early could represent mutants that are able to establish infection but unable to sustain long-term *in vivo* growth. Genes required early but not late could represent mutants that grow slowly in mice. These mutants would be underrepresented at d10, but catch up by d45 post infection. In fact, genes that were required at d10 but not at d45 were enriched for loss-of-insertions (Fig 3.3B). The average *in vitro*:*in vivo* ratio for these genes was 4.3, compared to the non-required average of 1.3 (P-value < 0.001) (Fig 3.3C).

While this set significantly overlapped with a previously defined set of genes required for growth in mice (Gene Set Enrichment Analysis P-val < 0.01), it also includes more than 400 newly discovered genes required for *in vivo* growth (Fig 3.4A). Despite the significant overlap, genes marked as essential in Sassetti et al. but not in this study did not enrich for high *in vitro*:*in vivo* ratio, unlike the non-overlapping sets between the two time points in this study. This is likely an indication of many different phenomena. The screens were performed with different libraries and mice were sacrificed and different time points. Both screens, as is true for most genome-wide studies, have high background noise, and many of the non-overlapping genes represent this noise.

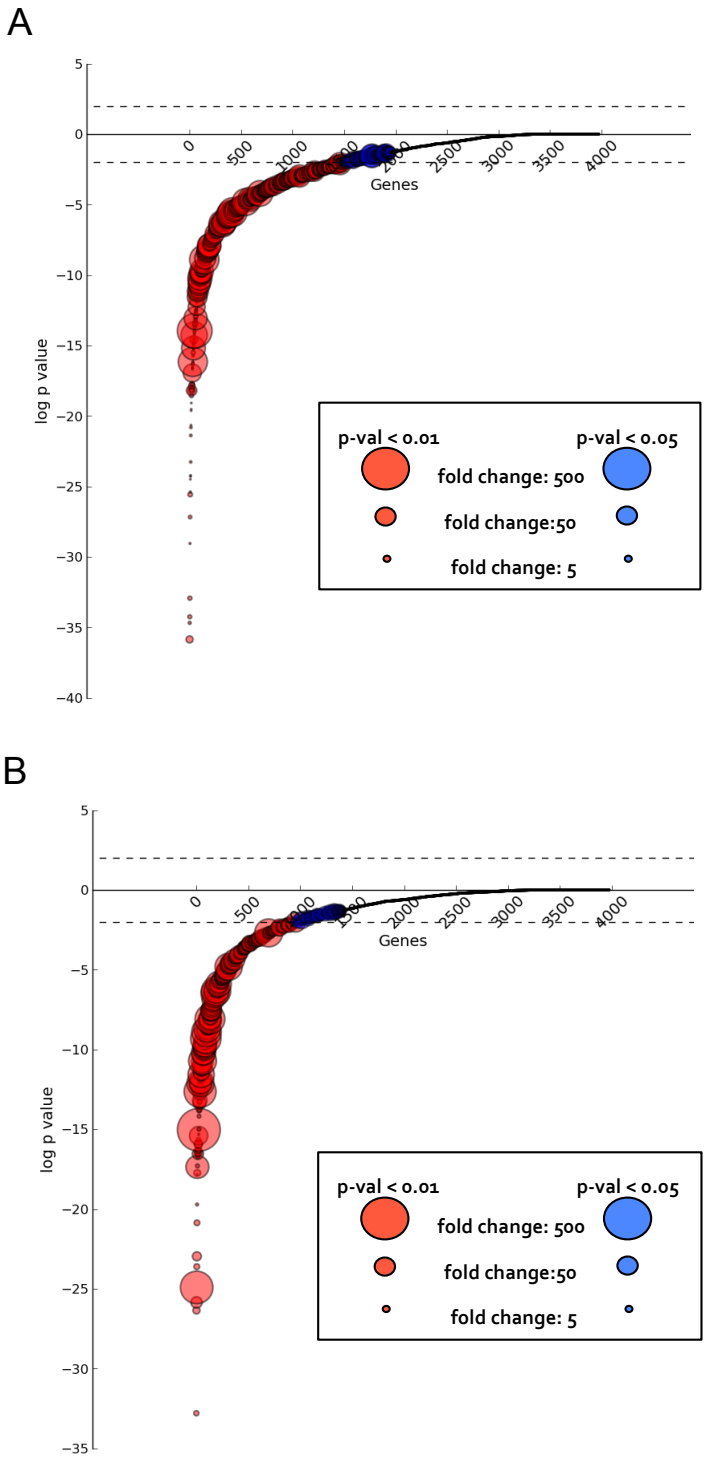


Figure 3.2. We compared insertions from wildtype mouse output libraries and MHCII$^{-/-}$ mouse libraries at d10 (A) and d45 (B) post-infection. After adjusting insertion counts for total reads in the sequencing runs, we calculated the ratio of insertions for each gene and calculated a P-value (Mann-Whitney U).

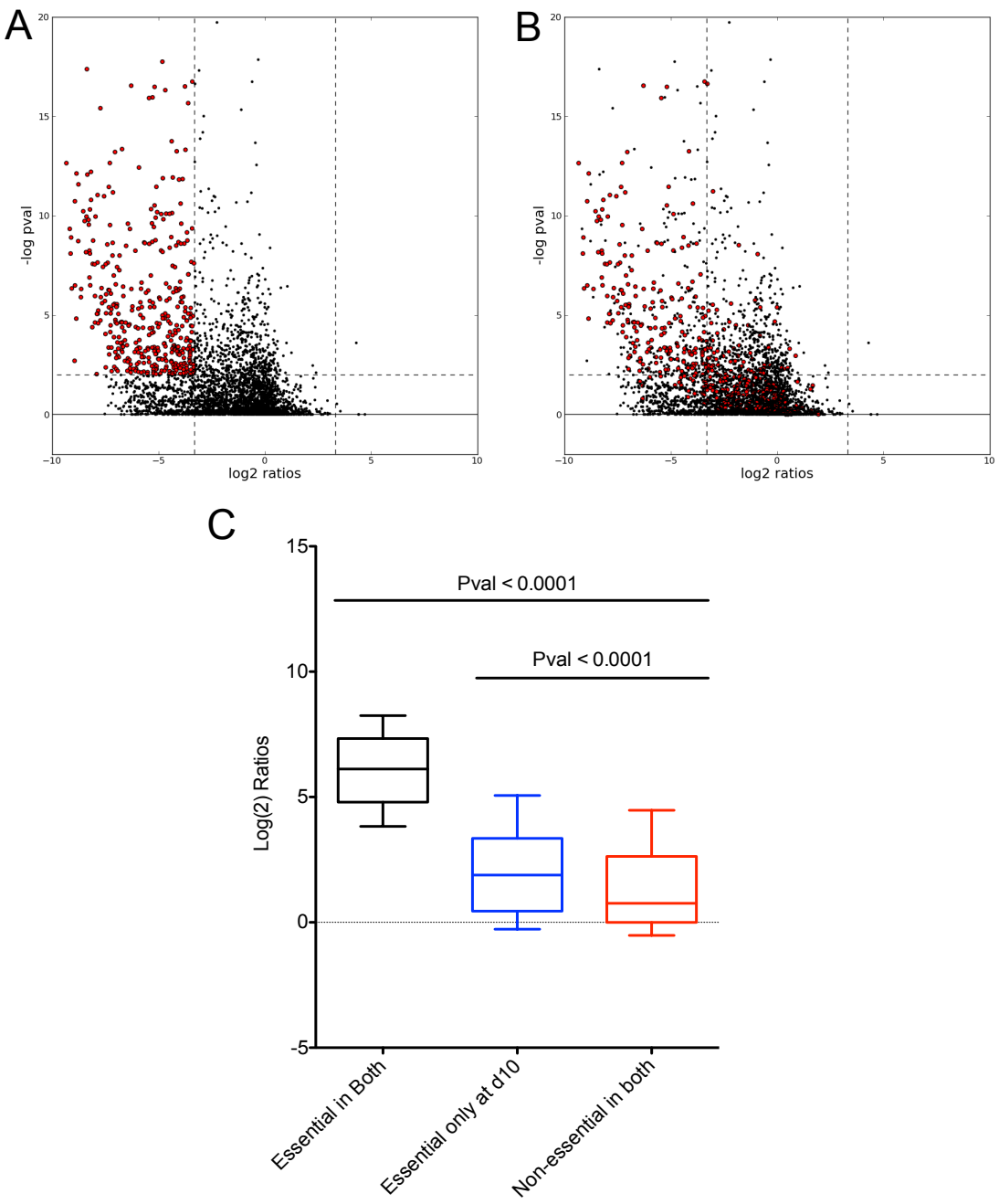
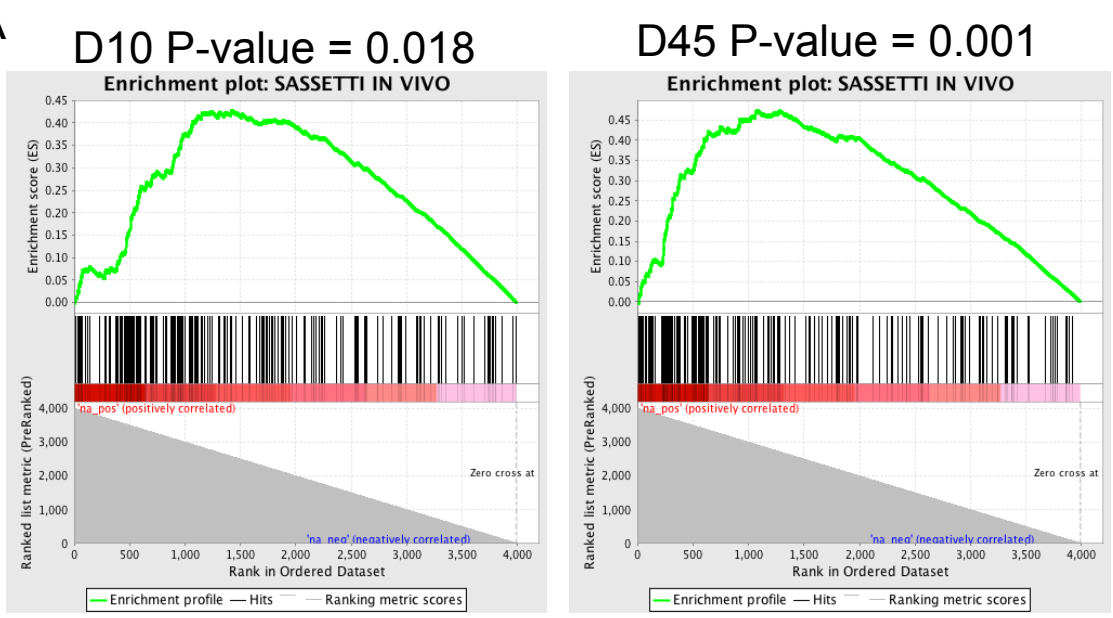


Figure 3.3. Log(2) ratios and Log(10) P-values of d45 wildtype mouse output libraries compared with inoculum libraries are plotted, each dot representing one gene. Red dots represent: A.) genes required for growth *in vivo* at d45 and B.) genes required for growth *in vivo* at d10. C.) Distributions of Log(2) ratios for genes tagged as essential for *in vivo* growth at both time points, essential only at day 10, and non-essential at both time points.

As discussed in Chapter 2, transposon mapping affords greater sensitivity compared to the microarray methods employed in Sassetti et al., and it is possible that many of the genes found only in this study represent the sensitivity space covered by deep sequencing and missed by microarrays[21]. Supporting this argument are the many metabolic pathways and known protein complexes, even those not encoded as an operon, wherein multiple components were tagged as required *in vivo*. These include the nucleotide excision repair pathway (*uvrA*, *uvrB*, *uvrC*); the glyoxylate shunt (*icl*, *glcB*); gluconeogenesis (*pca*, *pckA*); tryptophan biosynthesis (*trpD*, *trpE*, *trpB*); the bc cytochrome (*cydA, cydB, cydC, cydD*); mycobactin biosynthesis (*mbtA, mbtB, mbtD, mbtE, mbtF*); and the F420 synthetic pathway (*fbiA*, *fbiB*, *fbiC*), many of which have been shown to be required in other independent studies[22-26].

A



B

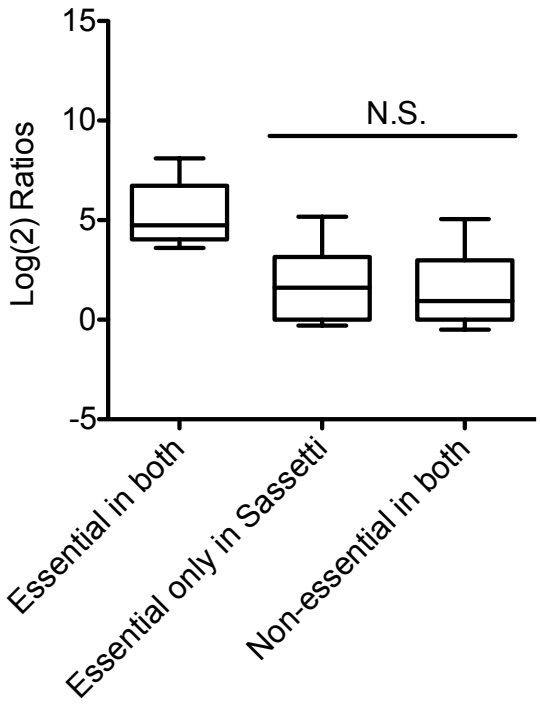


Figure 3.4. A.) Gene Set Enrichment Analysis (GSEA) using genes required in vivo in Sassetti et al. as a comparison gene set. For each time point, a list ranked by p-values and ratios was used to perform GSEA, and enrichment Family Wise Error Rate P-values were calculated. Enrichment plots represent running score determined by enrichment of the gene set in the ranked list. B.) Distributions of Log(2) ratios for genes tagged as essential in both studies, only in Sassetti et al., and non-essential in both studies.

Genes specifically required for surviving CD4-mediated stress

By comparing the wildtype output libraries to the MHCII$^{-/-}$ output libraries, we found that 58 genes had a statistically validated increase in insertions in MHCII$^{-/-}$ mice compared to wt mice. These genes had at least a 5-fold increase in insertions and a Mann-Whitney U P-value of less than 0.05 (Fig 3.5A and Table 3.2). Two biochemical pathways, gluconeogenesis and tryptophan biosynthesis, were enriched in this list, suggesting that CD4 T cells were responsible for inducing a shift in Mtb metabolic demands (Table 3.2).

Searching for gene sets that are enriched in high-throughput screen hit lists can help determine the biological functions uncovered in these screens. By looking for biochemical pathways enriched in our MHCII$^{-/-}$ rescue gene set, we determined that bacterial gluconeogenesis and tryptophan biosynthesis might be necessitated by CD4-dependent host defenses. We sought to further characterize CD4-mediated stress by more similar profiling, and we reasoned that conditionally required gene sets would be fitting comparison groups. To that end, we selected our Mtb transposon library through a series of *in vitro* stress conditions, expecting that conditions that most similarly resembled CD4-mediated stress would produce the most similar conditionally required gene sets.

Using the same transposon mapping technique on our selected library, we tested carbon starvation, amino acid starvation (and isolated tryptophan starvation), iron

depletion, acid stress and nitrosative stress. We created conditionally required gene sets as well as ranked gene lists for each condition, which allowed us to compare gene sets using the running-sum enrichment analysis in the Gene Set Enrichment Analysis (GSEA) tool[27]. Interestingly, while we found that both nitrosative stress and acid stress had significantly similar profiles to genes required for *in vivo* growth, only tryptophan starvation was enriched in the MHCII$^{-/-}$ rescue gene set (Fig 3.5B). Indeed, two genes in the tryptophan biosynthesis pathway, *trpD* and *trpE*, had no insertions in the wt output library but multiple insertions in the MHCII$^{-/-}$ output library (Figure 3.6). Tryptophan auxotrophs and their complemented strains have been generated and are currently being tested for growth in wildtype and MHCII$^{-/-}$ mice. We expect that growth of the auxotrophs will be severely limited in wildtype mice but restored in MHCII$^{-/-}$ mice.

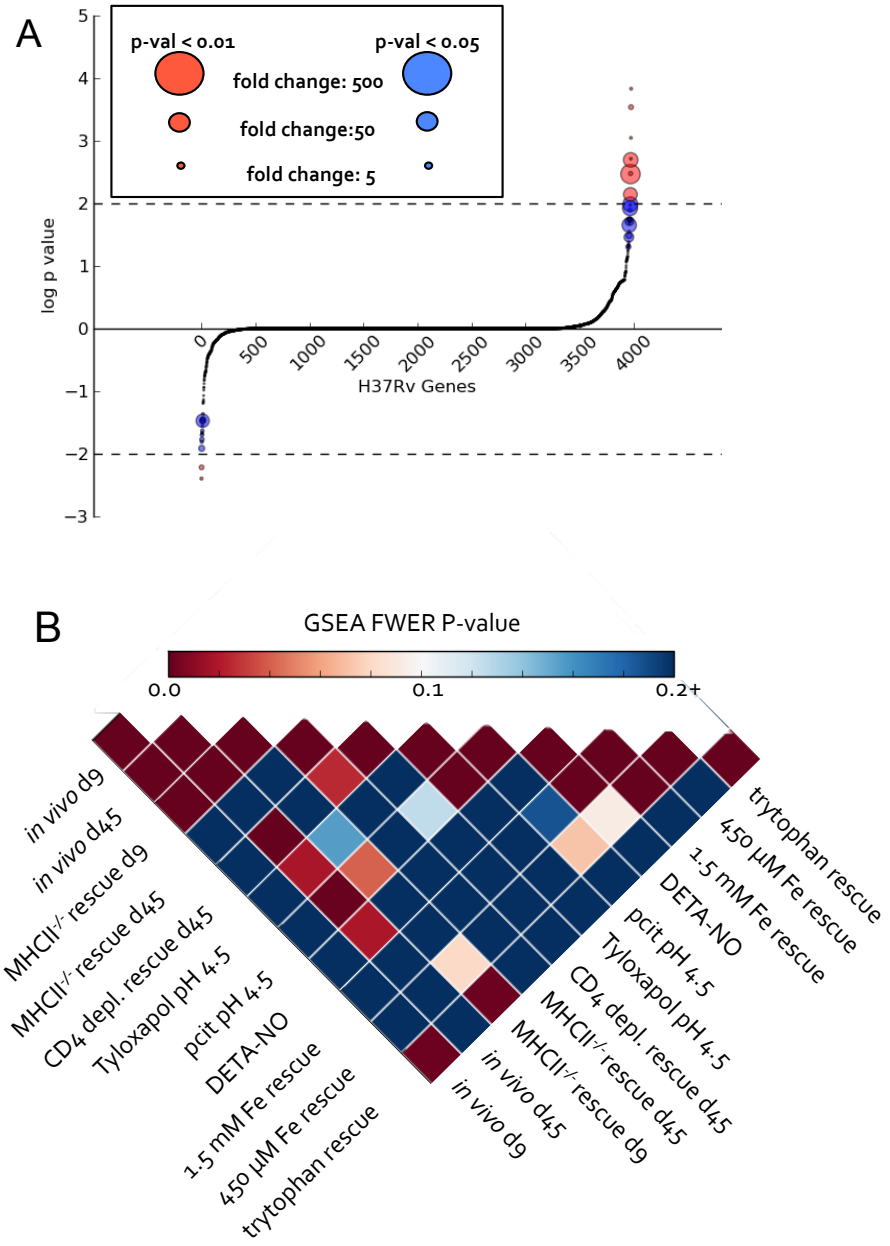


Figure 3.5. A.) Insertion counts ratios and P-values (Mann-Whitney U) were created for each gene between MHCII$^{-/-}$ and wt mice. P-values above the axis denote a MHCII$^{-/-}$: wildtype ratio > 1, and P-values below the axis denote a ratio < 1 for total-read-count-adjusted insertion counts. B.) GSEA Family-Wise Error Rate P-values for each pairwise comparison between libraries. Gene sets for each condition noted were used to perform GSEA on ranked lists from the same conditional essentiality experiments.

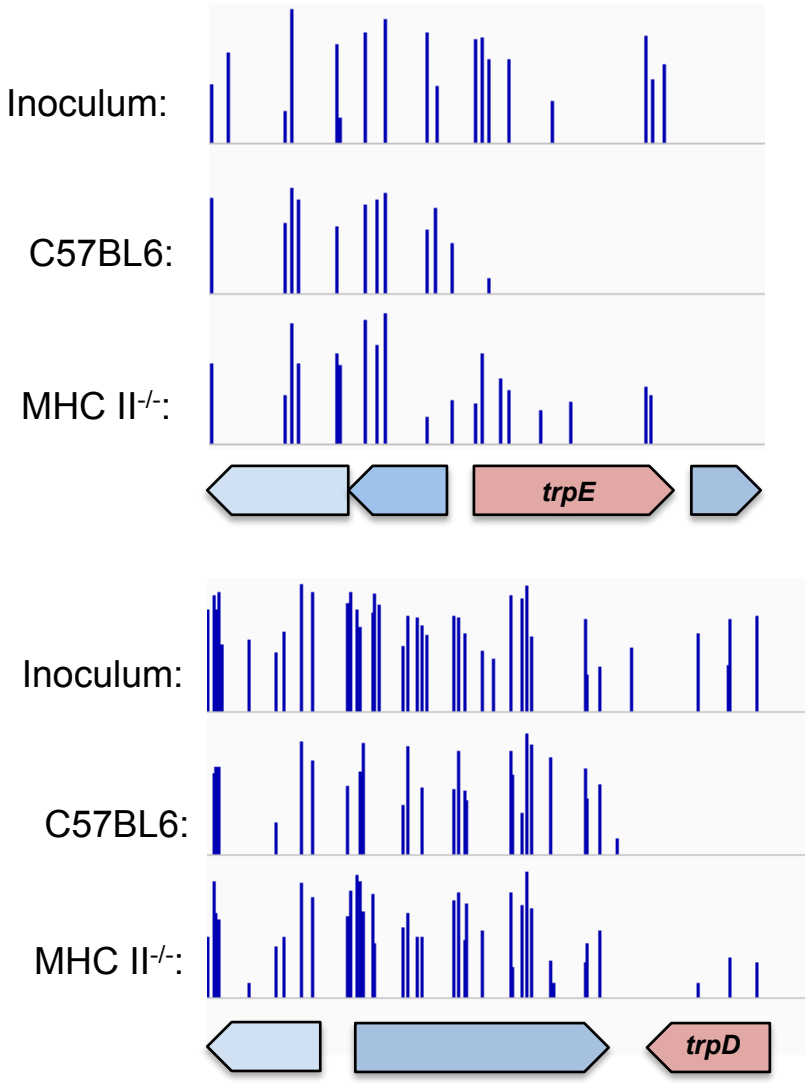


Figure 3.6. Insertion counts in the regions surrounding *trpE* and *trpD*. Both genes were required for growth *in vivo*, yet mutants in each gene were rescued for growth in MHCII$^{-/-}$ mice.

Discussion

CD4 T cells are paramount in the host defense against TB, but are insufficient to clear the bacteria from a diseased patient. We reasoned that genes required to survive the CD4 T cell response would be excellent drug targets. Inhibiting these CD4 survival determinants would restore CD4 T cell potency and help these cells clear the pathogen. Here we show that tryptophan biosynthesis is one of these determinants of survival, and inhibition of the pathway increases Mtb susceptibility to host immunity, ultimately leading to death of the pathogen *in vivo*. It is important to note that while the lack of CD4 T cells rescues growth of tryptophan auxotrophs early in infection, these mutants eventually do die in $MHCII^{-/-}$ mice, suggesting that compensatory CD4-independent mechanisms can also starve Mtb of tryptophan. Importantly, this shows that targeting tryptophan biosynthesis would be effective even in patients, such as HIV-positive individuals, who lack a normal CD4 compartment.

We started our study by profiling mycobacterial genes required for survival during infection. To determine the conditional requirement of genes, we compared the number of transposon insertions between conditions for each gene. Each gene has multiple potential insertion sites, and most insertion count comparisons sum counts across each gene to compare between conditions. However, in reasonably saturated libraries, treating each potential insertion site as an independent assessment of gene requirement drastically increased the statistical power of such comparisons. Most transposon insertion sequencing techniques

currently use insertion count totals, summed across insertion sites in a gene[28-31]. We found that this drastically decreases the power of statistical hypothesis testing, and increases the likelihood that a single differential site could make the entire gene look differentially required. Using a non-parametric test to compare the distribution of insertion counts (rather than the sums of insertion counts) across a gene, we were able to assess genetic requirement across a host of different conditions.

In contrast to similar microarray-based screens, deep sequencing allowed us to directly map insertion sites for each surviving mutant, and thus improved significantly on the sensitivity and specificity of insertion count comparisons. We found 576 genes that were required for growth *in vivo*, most of which are newly described as required. Furthermore, by searching for mutants rescued for growth in $MHCII^{-/-}$ mice, we described 58 genes that were likely required for surviving CD4-mediated host defenses. In addition to using these screening results to identify candidate genes, we also appreciated they could also be used as environmental profiles. We showed that the *in vivo* environment as a whole, as well as the specific CD4-generated component of that environment, could be understood by comparing multiple gene-requirement profiles. Using GSEA to compare *in vivo* profiles to *in vitro* profiles confirmed that Mtb's infectious niche is characterized by acidic and oxidative stress, and that CD4 T cells impose tryptophan starvation. We show that gene-requirement signatures are an effective way to understand both mycobacterial physiology as well as the

environments created by components of host immunity. Further profiling of *in vitro* stresses and altered *in vivo* environments will help us better understand the tug-of-war between host and pathogen.

In this study, high-throughput profiling also helped us narrow our search for candidate virulence factor drug targets. Targeting virulence factors has been touted recently for many reasons, including i.) expansion of the potential target list beyond *in vitro* essential genes, ii.) protection for the host against virulence factor toxicity, and iii.) reduction of resistance development for drugs whose targets do not affect viability[32, 33]. By searching for mycobacterial processes that protect Mtb from CD4 T cell-mediated killing, we hoped to identify drug targets whose inhibition made Mtb more susceptible to antimycobacterial immunity. Since tryptophan auxotrophs, which could not survive in wt mice, were rescued in the absence of CD4 T cells, we determined that blocking tryptophan biosynthesis would allow CD4 T cells to kill Mtb.

Methods and Materials

Library generation. Transposon libraries were created as described in Chapter 2, but plated on 7H10 plates with glycerol, OADC, tween80, Cas-amino acids and tryptophan. 500,000 mutants were scraped, frozen and saved for future use.

Mouse infections and harvests. Wildtype mice (C57BL/6) were obtained from Jackson Laboratory. MHC Class II knockout mice (Abb H2-Ab1) were obtained from Taconic Farms. These mice have a disruption in the H2-Ab gene, do not

express any MHC Class II molecules, and lack CD4 T cells. Mice were infected with 10^6 bacteria via tail vein injection. At 10 days and 45 days post-infection, spleens were harvested and plated for bacteria. For each mouse, 10^6 surviving colonies were scraped and DNA was extracted for analysis.

***In vitro* transposon library selections.** For acid stress, the library was suspended in liquid media at a starting concentration of 10^8 CFU/ml and selected in 7H10 with tyloxapol (in place of tween80) buffered to pH 6.5 or pH 4.5, and in phosphate-citrate buffer at pH 4.5. Bacteria was plated after 6 days and scraped for DNA prep. For nitrosative stress, the library was suspended in 5 mM DETA-NO for 3 days. Tryptophan starvation and amino acid starvation was measured using libraries that were created either on normal 7H10, or on 7H10 supplemented with 1 mM tryptophan or 1% Cas-amino acids, respectively. Iron supplemented libraries were plated on 7H10 with 450 μM and 1.5 mM iron.

DNA library prep and sequencing. See Chapter 2, Section 2.2 Methods and Materials.

Statistical Analysis. For each insertion count comparison, the control libraries were combined using a script that normalized insertion counts to the sequencing run's total read counts. This combined control library was then used to compare to each experimental library. For each gene, we treated the insertion counts at sites within the middle 90% of the gene as non-parametric distributions, and assumed the null hypothesis that the distributions would be the same between conditions. We used a Mann-Whitney U test for hypothesis testing. A P-value was thus calculated for each replicate, and a composite P-value was generated

by using a Bonferroni correction and Fisher's method. We then generated Benjamini-Hochberg false discovery rates using the composite P-values. Ratios were calculated by averaging the read counts per gene for all replicates and comparing to the combined control library, after normalizing for total read counts. Gene Set Enrichment Analysis was performed using the pre-ranked tool in GSEA. Genes were stratified by P-value (< 0.01 or 0.05 and > 0.01 or 0.05), and then ranked by ratio within the strata. For each conditional gene-requirement experiment, a ranked list and a conditional essential "calls" gene set was created. Each ranked list was then assessed against each gene set, and Family-wise Error Rate was used to generate the P-values for significant enrichment.

Section 3.3: Tables

Table 3.1. Genes required for growth *in vivo*

		Day 10			Day 45		
Rv Number	**Gene**	**MWU P-val**	**in vitro:in vivo ratio**	**FDR (q-val)**	**MWU P-val**	**in vitro:in vivo ratio**	**FDR (q-val)**
Rv0003	recF	2.75E-10	7.41E+01	8.82E-09	6.06E-06	9.34E+00	8.70E-05
Rv0007	-	4.08E-05	165.4828612	2.51E-04	3.27E-07	4.87E+02	7.26E-06
Rv0009	ppiA	0.002283889	32.50011905	7.53E-03	7.04E-05	22.03468221	0.000725051
Rv0012	-	7.02E-05	6.698676895	0.000401215	0.000567599	18.5988609	0.004086613
Rv0016c	pbpA	2.00E-12	1.47E+02	1.11E-10	2.26E-13	1.56E+02	2.47E-11
Rv0017c	rodA	5.53E-15	4.40E+02	4.78E-13	3.57E-12	3.43E+01	2.94E-10
Rv0018c	ppp	1.92E-05	45.07609164	1.35E-04	2.78E-08	1.79E+02	7.68E-07
Rv0019c	-	6.82E-06	1.04E+02	5.68E-05	0.000803865	14.46224068	0.005346102
Rv0021c	-	0.014070389	1.026198811	3.55E-02	2.68E-05	21.72475907	0.000316903
Rv0040c	mtc28	0.002210434	144.3599587	0.007325746	0.001705832	31.7316938	0.009991304
Rv0050	ponA1	2.69E-12	2.48E+02	1.43E-10	3.30E-05	2.565713435	0.000374744
Rv0057	-	0.000101612	128.6272784	0.000549813	0.01353601	4.037367438	0.052264768
Rv0066c	icd2	0.000343063	26.58650071	1.56E-03	8.07E-05	39.8296882	0.000802171
Rv0067c	-	4.68E-06	4.37E+01	4.11E-05	0.000101293	43.48519017	0.000973047
Rv0078	-	0.005494691	31.56679751	0.015743794	0.00076471	61.57866239	0.005137249
Rv0078A	-	0.000273864	22.9086043	0.001270805	0.700761503	0.501885082	1
Rv0086	hycQ	0.000769024	27.60259804	0.003034135	0.033906377	27.88025856	0.108049409
Rv0088	-	0.000995672	123.5393202	0.003746252	0.293015882	5.335606181	0.558264046
Rv0090	-	0.000522176	142.7371277	0.002197558	0.160992854	1.114289613	0.36173366
Rv0098	-	0.012198635	0.622890636	0.031319542	0.000614377	106.0472191	0.004345516
Rv0099	fadD10	6.81E-07	1.80E+00	7.76E-06	3.35E-08	2.02E+01	9.13E-07
Rv0120c	fusA2	2.63E-05	19.60701472	0.00017441	0.001127472	0.546650247	0.007050248
Rv0125	pepA	0.354404986	0.878989384	0.531542111	0.000198711	11.00532858	0.001729265
Rv0126	treS	1.68E-08	1.21E+00	3.34E-07	4.73E-11	4.08E+01	2.77E-09
Rv0127	-	0.000268335	66.94165164	1.25E-03	6.70E-05	97.57529103	0.000695936
Rv0155	pntAa	1.46E-07	1.56E+02	2.19E-06	2.41E-09	4.99E+01	8.55E-08
Rv0157	pntB	0.001096655	2.155590627	4.04E-03	4.65E-10	5.77E+02	2.01E-08
Rv0158	-	2.91E-06	1.18E+01	2.76E-05	0.053962132	0.336749401	0.150075105
Rv0162c	adhE1	2.72E-05	4.329033185	0.000180434	0.00019089	21.21771789	0.001679576
Rv0163	-	1.22E-06	2.42E+02	1.32E-05	0.080347229	1.562050066	0.207224988
Rv0177	-	1.98E-05	163.7601987	0.000138694	0.874087111	0.652052843	1
Rv0180c	-	1.09E-14	7.73E+02	8.49E-13	9.14E-16	1.16E+03	1.45E-13
Rv0184	-	0.000900199	124.3165166	0.003465723	0.085557982	2.154079144	0.218031397
Rv0188	-	0.00096369	112.7027378	0.003667381	0.066778806	1.596251034	0.178961801
Rv0199	-	0.000198556	220.5808513	0.000967718	0.003227811	9.317452657	0.016957733
Rv0204c	-	8.27E-08	1.68E+00	1.35E-06	8.65E-13	3.25E+02	8.19E-11
Rv0205	-	4.68E-07	2.09E+02	5.72E-06	0.032118785	1.125850373	0.103598061
Rv0207c	-	0.001020724	13.90234298	0.003829643	0.134663586	18.15535353	0.313191275
Rv0208c	trmB	0.025952798	39.80561504	5.95E-02	8.78E-06	118.2176481	0.000119948
Rv0211	pckA	4.69E-05	62.24241199	2.83E-04	4.69E-05	62.24241199	0.000502656
Rv0216	-	0.00105034	0.873481161	3.93E-03	5.12E-06	1.49E+02	7.54E-05
Rv0222	echA1	0.000605379	96.97821916	0.002469325	0.030769299	10.1190144	0.100385154
Rv0234c	gabD1	1.13E-09	2.00E+01	3.14E-08	0.000482442	4.100464405	0.003635004
Rv0235c	-	2.59E-07	2.25E+02	3.63E-06	0.006352202	2.384618234	0.028707623
Rv0238	-	0.001112296	120.8288942	0.004084582	0.001112296	120.8288942	0.006988313
Rv0241c	-	0.000496346	98.6854895	0.002115721	0.002587152	51.5530908	0.014133382
Rv0242c	fabG	8.78E-06	2.38E+02	7.03E-05	8.78E-06	238.2157481	0.000119948
Rv0243	fadA2	2.23E-07	1.33E+00	3.19E-06	8.12E-11	2.23E+01	4.25E-09
Rv0244c	fadE5	2.07E-14	5.91E+00	1.58E-12	4.30E-18	3.30E+02	1.55E-15
Rv0247c	-	8.28E-05	52.712884	0.000463176	0.000569269	35.98114003	0.004086613
Rv0248c	sdhA	1.44E-09	8.81E-01	3.82E-08	4.52E-14	1.05E+02	5.80E-12
Rv0249c	-	1.64E-10	1.83E+00	5.68E-09	1.53E-10	2.73E+01	7.51E-09
Rv0273c	-	9.89E-06	3.67E+01	7.82E-05	0.002231639	4.556535679	0.012500319
Rv0310c	-	0.019069064	1.314234537	0.04571288	0.000240422	121.9703665	0.002056252
Rv0312	-	4.01E-07	8.89E+01	5.19E-06	0.004134052	8.531905093	0.020680662
Rv0324	-	3.50E-06	5.30E+01	3.18E-05	1.02E-08	1.14E+02	2.98E-07

continued on next page

Rv0332	-	0.000397655	39.85350113	0.001765037	0.486375016	1.304701861	0.814447763
Rv0343	iniC	0.000347157	46.51695526	0.001572489	0.145393843	1.568392488	0.333274533
Rv0360c	-	0.000514951	18.72520201	0.002181002	0.003684332	9.347407664	0.01878537
Rv0361	-	4.11E-05	230.8451543	0.0002527	0.000185305	95.60089516	0.001634056
Rv0363c	fba	0.001816367	22.71285069	6.16E-03	1.87E-05	212.4761307	0.000234445
Rv0370c	-	1.94E-06	1.35E+01	1.99E-05	4.80E-06	6.17E+01	7.18E-05
Rv0380c	-	0.000868876	108.5586577	0.00337453	0.385100891	4.536104725	0.687147528
Rv0381c	-	6.32E-07	4.35E+00	7.26E-06	1.30E-05	14.04295565	0.000169782
Rv0384c	clpB	4.36E-05	52.98164753	2.67E-04	4.36E-05	52.98164753	0.000472597
Rv0392c	ndhA	3.50E-05	4.797618132	2.21E-04	5.93E-05	18.69234918	0.000620347
Rv0403c	mmpS1	0.000420961	100.2089866	0.001845824	0.066977975	1.211391861	0.179041117
Rv0407	fgd1	9.27E-09	1.02E+01	1.99E-07	5.15E-07	5.69E+00	1.04E-05
Rv0409	ackA	0.002351459	93.15167151	0.007709608	0.066961057	17.84988228	0.179041117
Rv0412c	-	7.84E-07	6.38E+01	8.74E-06	1.49E-05	27.81299966	0.000191144
Rv0432	sodC	3.88E-05	161.137982	0.00023995	0.002296804	2.983265128	0.012811206
Rv0436c	pssA	5.24E-09	2.29E+02	1.19E-07	1.90E-11	4.88E+02	1.26E-09
Rv0438c	moeA2	0.0003709	50.38692479	0.001664866	0.001086181	11.40040657	0.006835037
Rv0444c	-	0.000936537	47.16117285	0.003577914	0.017013112	2.907355283	0.0629406
Rv0465c	-	0.055962008	0.91248773	1.12E-01	1.21E-12	2.01E+01	1.12E-10
Rv0467	icl	3.35E-05	57.05272778	2.13E-04	2.82E-09	1.70E+02	9.49E-08
Rv0468	fadB2	0.000723572	16.35307407	0.002889201	0.000471952	9.78064932	0.003575151
Rv0470c	pcaA	3.47E-07	1.69E+01	4.61E-06	5.73E-09	2.94E+01	1.82E-07
Rv0472c	-	1.48E-06	5.83E+01	1.56E-05	0.000144362	12.65213846	0.001316804
Rv0485	-	8.78E-08	2.39E+00	1.43E-06	3.82E-13	6.05E+01	3.90E-11
Rv0489	gpm1	0.001369849	53.25323406	4.86E-03	5.69E-07	1.59E+02	1.14E-05
Rv0495c	-	2.63E-05	5.765104995	1.74E-04	9.09E-07	1.76E+02	1.71E-05
Rv0497	-	6.13E-06	1.42E+02	5.18E-05	4.21E-06	1.33E+01	6.34E-05
Rv0498	-	0.014906151	2.475370231	3.72E-02	1.05E-05	28.08630138	0.000140979
Rv0504c	-	1.98E-06	1.96E+02	2.02E-05	3.15E-06	1.40E+02	4.97E-05
Rv0505c	serB1	9.92E-05	87.47620658	0.000538246	0.000448575	32.84386754	0.003430733
Rv0519c	-	1.75E-05	62.70160318	0.000125043	0.064126643	2.62953397	0.172785678
Rv0541c	-	7.11E-05	81.01794719	4.05E-04	3.43E-05	84.73358477	0.000385271
Rv0545c	pitA	0.002723151	0.787559888	0.008740899	0.000400513	16.86132669	0.003147906
Rv0561c	-	4.91E-07	9.07E+00	5.87E-06	2.34E-09	1.42E+02	8.39E-08
Rv0563	htpX	0.000225098	5.659762643	1.08E-03	2.66E-05	124.436535	0.000316299
Rv0579	-	0.000143457	182.0799364	0.000737119	0.073494997	1.561316106	0.192295792
Rv0580c	-	0.000361546	48.37762624	0.001630235	0.053662934	5.086990607	0.149452024
Rv0582	-	0.003411848	7.973850628	0.01057593	0.000606287	17.18665648	0.004305724
Rv0586	-	0.000290157	72.40495939	0.001330973	0.386500105	2.68813958	0.68805323
Rv0588	yrbE2B	6.35E-05	125.4924525	0.000367254	0.149985326	7.145603081	0.341829021
Rv0590	mce2B	9.05E-05	49.90757681	0.000499699	0.001491491	9.488910364	0.008933223
Rv0590A	-	0.000191992	156.77054	0.000942659	0.023064459	5.829809438	0.080122354
Rv0619	galTb	0.000896658	76.18171628	0.003458788	0.004592089	18.19973709	0.022491059
Rv0622	-	2.45E-07	1.58E+02	3.47E-06	0.000651904	15.0401151	0.004564474
Rv0634A	-	0.000284239	205.0883153	0.001308355	0.090420896	14.67142185	0.228072538
Rv0642c	mmaA4	0.010820312	0.718342883	2.84E-02	2.54E-10	1.53E+01	1.18E-08
Rv0646c	lipG	0.000444536	193.5316048	0.001932151	0.037371033	3.669760468	0.115841463
Rv0671	lpqP	0.000449465	179.1352605	0.001949315	0.016674354	1.740765256	0.061859985
Rv0687	fabG	3.18E-05	26.92951262	0.000204734	0.018688614	6.234391804	0.068039921
Rv0694	lldD1	6.30E-05	89.87367635	0.000364512	0.369486649	0.535263766	0.665511053
Rv0695	-	0.00096978	15.46150479	0.003680167	0.370354013	4.647715607	0.666283893
Rv0712	-	2.67E-06	1.68E+01	2.58E-05	7.25E-05	10.81016869	0.000737568
Rv0725c	-	0.000892958	101.0100065	0.003447856	0.666636064	1.228500398	1
Rv0726c	-	0.002562459	166.1744342	0.008285284	0.616362248	6.917597116	0.965144631
Rv0727c	fucA	0.121338287	0.742390129	0.217370435	0.000333084	19.1331945	0.002720075
Rv0742	PE_PGRS 8	0.001610677	88.57933345	0.005546029	0.504014437	0.885658409	0.83103873
Rv0762c	-	0.001365421	1.671668775	0.004857008	0.000508626	12.44753532	0.003766864
Rv0764c	cyp51	0.000441403	112.9603049	0.001920634	0.658110484	0.331691244	0.999734682
Rv0784	-	0.000585995	21.20835836	0.002410032	0.008394253	14.18861244	0.035439433
Rv0798c	cfp29	0.001163778	12.83545369	0.004250088	0.005473068	1.757814685	0.025820154
Rv0805	-	7.85E-09	2.18E+02	1.71E-07	0.038598971	2.742748153	0.118447614
Rv0806c	cpsY	5.04E-14	2.43E+00	3.58E-12	1.36E-21	1.96E+01	7.71E-19
Rv0833	PE_PGRS 13	0.003987633	18.28226649	0.012078308	0.000411058	36.86344148	0.003200779
Rv0835	lpqQ	0.000712602	25.83296371	0.002853996	0.108751074	2.609726911	0.264366149

continued on next page

Rv0858c	-	5.52E-08	1.95E+02	9.26E-07	0.201093347	1.352853157	0.429510334
Rv0860	fadB	4.43E-07	8.87E+01	5.50E-06	1.02E-08	1.43E+02	2.98E-07
Rv0861c	ercc3	4.17E-08	1.36E+00	7.41E-07	2.09E-09	3.92E+01	7.85E-08
Rv0862c	-	1.11E-08	2.87E+01	2.35E-07	3.89E-06	1.71E+01	6.06E-05
Rv0867c	rpfA	3.07E-05	87.23068009	0.00019837	0.02227905	86.90146277	0.077812434
Rv0875c	-	3.69E-07	1.72E+02	4.86E-06	2.81E-08	2.56E+02	7.71E-07
Rv0876c	-	8.03E-12	1.58E+02	3.75E-10	3.62E-12	1.61E+02	2.94E-10
Rv0893c	-	2.16E-05	136.1949783	0.000148206	0.184790403	1.504521724	0.405695946
Rv0896	gltA	1.26E-09	5.52E+02	3.49E-08	1.26E-09	5.52E+02	5.07E-08
Rv0907	-	5.44E-08	1.20E+02	9.17E-07	0.001174723	5.776100777	0.007288415
Rv0910	-	0.001448861	130.428645	0.00508572	0.001448861	126.7367803	0.00870411
Rv0918	-	0.000433211	65.07559223	0.001893275	0.035199345	0.312955745	0.11095679
Rv0923c	-	0.004516974	0.695026615	0.013415987	0.000142739	15.46469351	0.001307999
Rv0926c	-	1.71E-08	1.14E+02	3.38E-07	8.32E-05	7.027670586	0.000818625
Rv0928	pstS3	2.76E-08	5.09E+01	5.27E-07	0.000500351	0.664692361	0.003712489
Rv0929	pstC2	4.16E-09	2.74E+02	9.73E-08	0.000114053	1.747709843	0.001074854
Rv0940c	-	0.000193404	120.1858586	0.000948418	0.03524559	22.64126717	0.11095679
Rv0946c	pgi	0.000167262	168.7991406	0.000838841	0.000259959	110.1212401	0.00217654
Rv0949	uvrD1	5.20E-12	3.25E+01	2.59E-10	3.08E-11	3.64E+01	1.92E-09
Rv0950c	-	6.99E-12	2.80E+02	3.43E-10	8.37E-11	2.93E+01	4.32E-09
Rv0955	-	1.27E-05	1.07E+00	9.50E-05	6.67E-11	3.34E+01	3.68E-09
Rv0959	-	0.000596188	66.47301885	0.002439343	0.903012782	0.810511296	1
Rv0966c	-	1.91E-05	123.984868	0.000134275	0.00284157	14.23774336	0.015266421
Rv0972c	fadE12	0.002817522	18.16702756	0.009000228	0.206187067	3.69820649	0.435247328
Rv0973c	accA2	4.69E-05	41.8624206	0.000283412	0.012559386	2.097985439	0.049113745
Rv0976c	-	9.13E-06	2.47E+01	7.26E-05	0.003343513	1.517111467	0.01747776
Rv0981	mprA	0.000257308	2.204648604	1.21E-03	4.00E-06	5.51E+01	6.14E-05
Rv0983	pepD	2.07E-06	2.04E+01	2.08E-05	2.63E-07	2.50E+01	5.94E-06
Rv0986	-	3.54E-05	173.6244585	0.000222955	0.01448253	5.494997494	0.055275453
Rv0988	-	6.81E-05	153.8500948	0.00039193	0.020529093	15.3964972	0.073157888
Rv0994	moeA1	5.69E-06	2.09E+02	4.90E-05	4.81E-07	5.65E+01	9.91E-06
Rv0998	-	0.014139765	28.32611366	3.56E-02	9.01E-05	180.1667934	0.000880521
Rv1010	ksgA	1.18E-05	5.44E+01	8.93E-05	0.000755805	20.74368577	0.005094636
Rv1028c	kdpD	1.40E-09	3.89E+01	3.76E-08	1.84E-06	2.05E+00	3.22E-05
Rv1050	-	0.00249897	89.42252564	0.008126252	0.40133088	1.868675596	0.706861341
Rv1058	fadD14	1.60E-06	9.22E-01	1.66E-05	2.44E-10	3.96E+01	1.16E-08
Rv1059	-	3.02E-05	45.63720189	1.96E-04	1.41E-05	52.9199517	0.00018087
Rv1065	-	0.287640083	0.769698744	4.46E-01	7.23E-07	1.98E+01	1.42E-05
Rv1078	pra	3.32E-05	152.4491925	2.12E-04	3.32E-05	152.4491925	0.00037548
Rv1080c	greA	0.003097682	26.45324392	0.009769614	0.000434348	39.38121526	0.00333475
Rv1082	mca	0.0005195	40.22863497	0.002188615	0.077242469	3.875222083	0.200680849
Rv1086	-	1.39E-06	2.25E+02	1.49E-05	8.10E-09	5.61E+02	2.42E-07
Rv1096	-	2.51E-07	2.38E+02	3.54E-06	0.000378644	25.50075559	0.002999733
Rv1097c	-	5.94E-07	2.00E+01	6.88E-06	0.000437491	1.320897681	0.003352413
Rv1099c	glpX	0.001074004	1.697939315	3.98E-03	5.95E-06	2.28E+02	8.60E-05
Rv1100	-	9.57E-05	172.8508281	0.000523757	0.257301044	1.934364361	0.503585754
Rv1105	-	6.16E-05	421.3731198	0.000357641	0.445123998	3.666284459	0.761401351
Rv1111c	-	0.001899535	198.3329164	0.006418396	0.026378372	3.297653292	0.088379769
Rv1112	-	2.27E-06	6.27E+01	2.24E-05	4.03E-06	4.21E+01	6.16E-05
Rv1121	zwf1	3.59E-10	1.17E+02	1.13E-08	0.004589606	0.708422188	0.022491059
Rv1127c	ppdK	0.008496012	0.513299159	2.29E-02	2.79E-06	1.35E+02	4.52E-05
Rv1128c	-	2.72E-10	1.40E+02	8.78E-09	4.09E-06	1.00E+00	6.21E-05
Rv1130	-	4.05E-12	2.60E+00	2.04E-10	4.42E-10	1.07E+01	1.93E-08
Rv1131	gltA1	1.15E-05	4.22E+01	8.83E-05	4.68E-05	42.56289383	0.000502656
Rv1144	-	0.000175375	17.53073343	0.000871832	0.692783928	1.242600502	1
Rv1159	pimE	6.85E-17	5.43E+02	8.25E-15	2.92E-17	7.79E+01	7.26E-15
Rv1170	mshB	5.69E-07	3.69E+02	6.64E-06	0.000129754	136.6784709	0.001197285
Rv1173	fbiC	1.29E-09	4.65E+01	3.53E-08	2.29E-13	6.44E+02	2.47E-11
Rv1178	-	0.003010514	1.415294442	9.53E-03	2.74E-05	30.81916164	0.000321583
Rv1187	rocA	0.000243532	117.5299463	0.00115991	0.437098773	6.25302775	0.750255425
Rv1193	fadD36	2.92E-07	2.00E+02	4.01E-06	4.44E-07	2.05E+02	9.35E-06
Rv1220c	-	0.001325914	150.1838947	0.004737529	0.000646962	6.985547508	0.004545879
Rv1221	sigE	0.001425758	70.75179174	0.005026809	0.020829827	2.514868047	0.073832638
Rv1234	-	0.000566159	106.8842643	0.002347877	0.194731094	3.506406004	0.421352318
Rv1235	lpqY	8.10E-12	1.61E+01	3.75E-10	3.06E-09	3.48E+00	1.02E-07
Rv1236	sugA	1.43E-05	21.52972219	1.05E-04	3.04E-06	8.14E+00	4.83E-05

continued on next page

Rv1238	sugC	7.04E-05	6.583728987	0.00040145	0.000120739	23.49807393	0.001129837
Rv1255c	-	0.000506864	11.05865995	0.002151332	0.041978888	10.21628783	0.125442379
Rv1260	-	0.000446149	90.17837731	0.001937047	0.754435029	0.705452271	1
Rv1272c	-	2.11E-09	4.65E+00	5.27E-08	7.44E-11	2.05E+01	4.00E-09
Rv1273c	-	3.00E-11	3.24E+02	1.22E-09	2.48E-11	1.56E+01	1.57E-09
Rv1278	-	3.11E-06	1.03E+02	2.90E-05	0.005070314	1.549616001	0.024265511
Rv1282c	oppC	7.72E-07	2.21E+01	8.64E-06	0.002065186	3.909876582	0.011683138
Rv1285	cysD	0.000127653	51.34844158	0.000668873	0.000658042	36.62947199	0.00458814
Rv1286	cysN	1.29E-06	1.18E+02	1.38E-05	1.34E-06	3.12E+01	2.44E-05
Rv1287	-	0.000757372	182.6118056	0.003000068	0.034504029	7.998162454	0.109253602
Rv1298	rpmE	0.000272518	48.45318936	0.001267606	0.000921524	34.56414037	0.006027799
Rv1322A	-	0.002476215	154.1302524	0.008058843	0.013184584	8.865515899	0.051106327
Rv1325c	PE_PGRS 24	0.000153331	81.92139492	0.000781792	0.005029715	66.39236647	0.024114524
Rv1333	-	0.000158758	1.907719682	0.000805334	0.000279889	26.78895918	0.002328701
Rv1337	-	2.72E-06	3.27E+02	2.61E-05	5.50E-06	7.13E+00	8.07E-05
Rv1339	-	2.23E-06	2.26E+00	2.21E-05	6.29E-07	1.79E+01	1.24E-05
Rv1340	rph	2.36E-05	71.03403806	1.60E-04	2.88E-06	1.24E+02	4.60E-05
Rv1346	fadE14	3.04E-05	119.5263715	0.000196952	0.15521003	0.773174372	0.350124952
Rv1348	-	0.006069997	1.05278722	1.72E-02	8.28E-05	18.38653974	0.000817608
Rv1349	-	2.96E-05	82.30513437	1.93E-04	2.96E-05	82.30513437	0.000340608
Rv1350	fabG	0.001627603	77.95257209	0.005589792	0.973463793	0.631569542	1
Rv1353c	-	0.000104026	112.4915833	0.000561346	0.096117137	11.18557521	0.239960987
Rv1355c	moeY	5.44E-13	1.84E+02	3.09E-11	2.55E-05	8.552522872	0.000304958
Rv1361c	PPE19	0.00175089	11.45422693	0.005966828	0.284258259	2.504790942	0.545406378
Rv1372	-	1.05E-05	8.48E+01	8.24E-05	0.177411657	0.902159183	0.393292174
Rv1373	-	2.09E-08	1.48E+01	4.06E-07	0.003475798	1.250882184	0.017975963
Rv1377c	-	1.42E-05	85.60110447	0.000104141	0.02740257	2.635497478	0.091120419
Rv1400c	lipI	1.15E-05	5.18E+01	8.83E-05	0.000802658	7.835866081	0.005346102
Rv1405c	-	0.000463534	15.50688517	0.002005956	0.00088063	11.43197552	0.005798454
Rv1410c	-	5.08E-13	2.63E+00	2.93E-11	1.82E-14	2.08E+01	2.50E-12
Rv1411c	lprG	0.00483274	92.94196902	1.42E-02	4.11E-05	276.8804084	0.000447936
Rv1420	uvrC	5.35E-08	1.28E+02	9.13E-07	1.13E-05	10.39124982	0.000150298
Rv1421	-	0.006686126	0.508898409	1.86E-02	3.87E-08	1.23E+02	1.04E-06
Rv1430	PE16	1.37E-18	2.51E+01	2.38E-16	8.66E-09	1.88E+00	2.57E-07
Rv1432	-	1.02E-09	3.46E+01	2.91E-08	1.06E-11	1.89E+02	7.52E-10
Rv1441c	PE_PGRS 26	2.86E-08	1.37E+02	5.44E-07	3.34E-05	3.837722097	0.000375948
Rv1448c	tal	0.000138405	124.1233542	0.000714851	0.002343591	20.98614209	0.012999248
Rv1458c	-	8.09E-06	1.04E+02	6.57E-05	1.30E-07	3.03E+02	3.18E-06
Rv1473	-	2.80E-07	4.42E+00	3.88E-06	2.86E-10	3.10E+02	1.31E-08
Rv1474c	-	1.89E-05	174.5524349	0.000133575	0.033238357	2.234500009	0.106261211
Rv1490	-	2.53E-07	1.13E+00	3.56E-06	6.59E-06	1.34E+01	9.40E-05
Rv1492	mutA	0.018948963	3.695171076	0.0454798	0.001440666	13.98364229	0.008681104
Rv1507A	-	0.001565659	83.54498737	0.005419169	0.474286771	5.408470316	0.798576836
Rv1507c	-	4.40E-05	101.6662035	0.000269509	0.058660718	5.949457348	0.160892189
Rv1509	-	0.0005433	110.772739	0.002274425	0.08704672	11.43586434	0.221345783
Rv1512	epiA	1.27E-06	2.10E+02	1.37E-05	1.27E-06	2.10E+02	2.33E-05
Rv1530	adh	1.33E-05	1.59E+00	9.85E-05	0.000210687	11.1819638	0.001809727
Rv1533	-	0.000627625	82.75201782	0.002547003	0.288551744	4.922238947	0.551716483
Rv1534	-	0.001309805	1.426053119	0.00469288	0.000347703	19.58657999	0.002804902
Rv1543	-	0.001733072	1.800649585	5.92E-03	6.68E-09	2.80E+01	2.06E-07
Rv1544	-	0.00063346	3.226201397	2.57E-03	2.43E-06	3.84E+01	4.05E-05
Rv1552	frdA	9.11E-10	2.31E+01	2.62E-08	0.005214725	0.80721244	0.024807369
Rv1559	ilvA	9.01E-05	0.843766208	4.98E-04	6.28E-13	2.89E+02	6.25E-11
Rv1565c	-	6.34E-19	3.40E+01	1.20E-16	2.36E-17	9.72E+00	6.25E-15
Rv1566c	-	6.02E-06	1.36E+02	5.11E-05	8.39E-07	1.51E+02	1.60E-05
Rv1568	bioA	2.19E-07	2.59E+01	3.15E-06	6.92E-09	3.40E+02	2.10E-07
Rv1569	bioF1	0.000105653	71.91847732	5.68E-04	2.40E-05	105.1965744	0.000289627
Rv1589	bioB	1.48E-06	2.35E+02	1.56E-05	1.03E-07	1.37E+02	2.56E-06
Rv1592c	-	1.46E-05	1.975708101	1.07E-04	1.27E-10	3.51E+01	6.34E-09
Rv1604	impA	0.000995672	156.4735211	0.003746252	0.087610469	10.89264733	0.222352799
Rv1609	trpE	2.32E-05	39.00349371	1.58E-04	1.11E-06	1.01E+02	2.07E-05
Rv1620c	cydC	0.026461397	14.60546331	0.060307722	0.00073058	130.5601983	0.00498444
Rv1621c	cydD	6.00E-08	1.80E+01	9.99E-07	1.62E-11	2.88E+02	1.11E-09
Rv1622c	cydB	5.22E-09	2.75E+02	1.19E-07	2.35E-08	1.68E+02	6.62E-07

continued on next page

Rv1623c	cydA	6.85E-07	1.43E+02	7.79E-06	4.70E-10	8.12E+01	2.01E-08
Rv1626	-	0.002532833	33.45778098	0.00820283	0.014199242	1.9892934	0.054335835
Rv1627c	-	3.82E-05	2.066446924	2.37E-04	3.47E-05	11.06678216	0.000389046
Rv1633	uvrB	1.48E-18	1.70E+00	2.45E-16	2.41E-24	2.03E+01	1.92E-21
Rv1636	TB15.3	2.40E-05	253.3424297	1.61E-04	2.40E-05	253.3424297	0.000289627
Rv1638	uvrA	6.56E-16	3.58E+02	6.69E-14	3.35E-17	3.63E+01	7.40E-15
Rv1660	pks10	0.000714966	178.7949622	0.002860583	0.128040919	1.048864719	0.301491257
Rv1663	pks17	2.48E-05	29.91165208	0.00016552	0.009373891	7.066442191	0.038955033
Rv1664	pks9	1.07E-14	6.27E+00	8.49E-13	7.99E-11	2.49E+01	4.24E-09
Rv1665	pks11	2.18E-07	2.04E+01	3.15E-06	0.003897491	3.164267179	0.019745634
Rv1683	-	5.37E-10	2.02E+01	1.63E-08	2.67E-06	4.49E+00	4.38E-05
Rv1688	mpg	0.0012296	105.6867972	0.004449607	0.488866191	1.382920427	0.816417255
Rv1700	-	0.012390218	53.13799766	0.031697488	0.000241846	104.8099957	0.002063995
Rv1701	xerD	0.000171187	76.41493922	0.000853146	0.001604793	32.26030442	0.009511568
Rv1710	-	0.001483132	11.4846308	0.005183142	0.021899888	1.42284667	0.07688163
Rv1732c	-	0.000881018	86.78297245	3.41E-03	8.19E-06	176.4966541	0.000115984
Rv1737c	narK2	1.08E-07	8.83E+01	1.68E-06	0.227708393	2.443722076	0.464885154
Rv1753c	PPE24	7.74E-12	2.33E+02	3.70E-10	0.045184169	1.046972312	0.132130471
Rv1755c	plcD	0.000284239	81.94223301	0.001308355	0.109135226	3.008526542	0.265137932
Rv1770	-	0.000161998	18.99106344	0.000818745	0.433899765	1.073596392	0.746375158
Rv1775	-	2.58E-06	9.13E+01	2.50E-05	0.003344374	6.840800655	0.01747776
Rv1791	PE19	0.001406876	50.7165919	0.004974495	0.001306523	51.74632518	0.007993912
Rv1809	PPE33	0.000731983	2.025702276	2.91E-03	1.31E-06	1.51E+01	2.38E-05
Rv1821	secA2	3.09E-16	1.05E+01	3.42E-14	1.81E-18	2.82E+01	7.19E-16
Rv1823	-	0.000679145	15.16134321	0.002739311	0.002874935	15.42203418	0.015388445
Rv1825	-	0.000720664	141.3189682	0.002880484	0.000727629	11.36242128	0.00498444
Rv1829	-	2.92E-06	5.33E+02	2.76E-05	4.62E-07	5.40E+02	9.62E-06
Rv1836c	-	5.61E-14	1.33E+00	3.92E-12	1.39E-26	8.01E+01	1.84E-23
Rv1837c	glcB	1.12E-10	2.49E+02	4.15E-09	1.12E-10	2.49E+02	5.62E-09
Rv1845c	-	2.06E-08	1.35E+01	4.01E-07	4.91E-11	2.99E+02	2.83E-09
Rv1854c	ndh	3.22E-07	2.96E+02	4.34E-06	3.22E-07	2.96E+02	7.20E-06
Rv1860	apa	7.82E-05	3.967419303	4.40E-04	3.89E-08	1.87E+01	1.04E-06
Rv1908c	katG	7.80E-11	3.36E+00	3.01E-09	2.12E-08	1.10E+01	6.07E-07
Rv1911c	lppC	0.000243923	71.09541602	0.001160384	0.367344364	1.06464211	0.662853237
Rv1913	-	2.50E-06	9.88E+01	2.43E-05	0.040177843	3.993810982	0.121975023
Rv1926c	mpt63	0.003092053	85.60336653	0.009769614	0.608758402	0.631055109	0.956551626
Rv1928c	-	1.37E-05	12.6317215	0.000100855	0.057567279	5.03752409	0.158879299
Rv1933c	fadE18	0.002253145	66.22956157	0.007442488	0.541827913	2.00709469	0.872897461
Rv1961	-	0.000821578	20.16144658	0.0032128	0.006662451	4.497207509	0.02964395
Rv1962c	-	5.76E-05	149.3736544	0.000339903	0.013176991	2.074861764	0.051106327
Rv1963c	mce3R	2.75E-06	1.58E+02	2.63E-05	2.31E-05	109.2228665	0.000282216
Rv2001	-	0.000516202	121.1763807	0.002183975	0.931796493	0.760246963	1
Rv2004c	-	1.05E-05	1.29E+02	8.22E-05	0.001641507	25.80962641	0.009700259
Rv2038c	-	0.001602828	2.744909149	0.005523784	0.000330702	89.54644217	0.002706179
Rv2039c	-	0.000484628	194.3968051	0.002074669	0.003936373	15.0968211	0.019891938
Rv2040c	-	0.000615561	18.94415269	0.002503155	0.002811632	11.23199306	0.015172133
Rv2043c	pncA	0.000668639	89.07507208	0.002702416	0.083635464	4.17052283	0.214040051
Rv2047c	-	1.84E-16	2.47E+00	2.09E-14	1.54E-12	1.64E+01	1.33E-10
Rv2048c	pks12	8.86E-30	1.95E+00	7.05E-27	1.56E-33	1.59E+01	6.22E-30
Rv2051c	ppm1	4.42E-07	4.57E+02	5.50E-06	2.87E-07	1.29E+02	6.45E-06
Rv2069	sigC	0.001792967	23.19853026	0.006089351	0.000700426	4.834467417	0.004836103
Rv2072c	cobL	0.000153988	11.78700013	0.000784134	0.471092288	0.450552649	0.794543694
Rv2079	-	7.02E-05	15.97554886	0.000401215	0.500126193	0.778050664	0.827171168
Rv2089c	pepE	0.000186777	129.5060986	0.000921216	0.004839515	25.10083592	0.023414539
Rv2091c	-	4.98E-06	9.46E-01	4.33E-05	3.96E-16	2.12E+02	6.84E-14
Rv2096c	-	0.000133285	68.27614707	0.000695635	0.132106175	4.346675784	0.308325269
Rv2097c	-	0.110762482	2.345667661	2.02E-01	4.02E-05	66.97517656	0.000440829
Rv2115c	-	6.39E-06	1.69E+02	5.36E-05	1.11E-06	2.54E+02	2.07E-05
Rv2119	-	6.78E-09	2.36E+02	1.50E-07	0.004578154	0.864114527	0.022478171
Rv2131c	cysQ	0.000731937	69.33856786	0.002911097	0.000731937	69.33856786	0.00498444
Rv2138	lppL	3.91E-06	1.26E+02	3.52E-05	3.91E-06	1.26E+02	6.08E-05
Rv2140c	TB18.6	0.000457509	51.13642252	0.001982042	0.00018178	59.64756297	0.001610108
Rv2170	-	0.000589246	80.46060723	2.42E-03	5.49E-07	1.63E+02	1.10E-05
Rv2192c	trpD	0.000434348	61.13059981	0.001896158	0.000434348	61.13059981	0.00333475
Rv2198c	mmpS3	8.10E-09	2.99E+02	1.75E-07	8.10E-09	2.99E+02	2.42E-07
Rv2206	-	0.003470546	30.34111481	1.07E-02	1.95E-05	12.70227707	0.000241925

continued on next page

Rv2207	cobT	2.45E-05	122.5683263	1.64E-04	1.98E-06	1.83E+02	3.40E-05
Rv2208	cobS	0.001069185	52.62478273	0.003973969	0.051563969	1.39112482	0.145028222
Rv2222c	glnA2	1.29E-10	2.57E+02	4.67E-09	2.29E-08	2.33E+02	6.50E-07
Rv2224c	-	1.14E-13	1.16E+01	7.43E-12	1.83E-17	1.06E+01	5.24E-15
Rv2230c	-	1.88E-05	88.66073421	0.000133323	0.000551054	10.41662387	0.003999163
Rv2231c	cobC	4.54E-06	1.11E+00	4.00E-05	2.55E-10	8.67E+01	1.18E-08
Rv2237	-	3.30E-06	2.72E+02	3.03E-05	0.053051461	3.591214684	0.148164087
Rv2241	aceE	7.15E-17	1.33E+00	8.37E-15	2.21E-16	1.22E+01	3.99E-14
Rv2249c	glpD1	0.000819463	58.93026435	0.003207682	0.073368503	1.802336399	0.192091202
Rv2250c	-	0.000995221	90.96801097	0.003746252	1	0.74993255	1
Rv2252	-	0.002241411	101.6328517	0.007416049	0.401778309	1.017408349	0.707336138
Rv2253	-	2.09E-11	2.99E+02	8.92E-10	1.21E-09	2.11E+01	4.90E-08
Rv2257c	-	0.035030629	54.63778174	0.076547698	0.000434348	162.630375	0.00333475
Rv2259	adhE2	4.62E-07	2.60E+00	5.70E-06	1.73E-09	3.52E+01	6.75E-08
Rv2269c	-	0.000359438	137.4931591	0.00162257	0.211707454	5.222696675	0.441279111
Rv2274c	-	2.40E-05	143.1038523	0.000161231	0.024004683	10.34236828	0.082440955
Rv2275	-	0.000438823	35.35557142	0.0019115	0.011914308	5.106389203	0.047007145
Rv2276	cyp121	2.44E-06	1.73E+02	2.38E-05	0.002240606	87.21036665	0.012532898
Rv2305	-	0.000213717	99.35509301	0.001030245	0.128793902	1.723729984	0.302726566
Rv2307B	-	0.00018518	98.18974584	0.000915998	0.045601471	8.786184171	0.132667923
Rv2317	uspB	0.000505698	110.1649546	0.002148678	0.0005303	65.14092324	0.003891147
Rv2322c	rocD1	0.000427507	46.01814771	0.001870404	0.018753026	2.567974576	0.068048161
Rv2334	cysK1	0.000855225	10.73994793	0.00332476	0.027361201	5.284950801	0.091058994
Rv2358	-	0.002054127	66.22072435	0.006836204	0.007496105	45.53081346	0.032510368
Rv2363	amiA2	7.56E-05	1.652506013	4.28E-04	1.92E-09	4.38E+02	7.29E-08
Rv2374c	hrcA	0.000372337	1.199430649	1.67E-03	1.32E-12	2.78E+01	1.19E-10
Rv2379c	mbtF	2.10E-10	1.14E+02	7.03E-09	9.24E-12	2.33E+02	6.81E-10
Rv2380c	mbtE	6.72E-10	1.22E+02	1.98E-08	6.34E-14	1.32E+02	7.41E-12
Rv2381c	mbtD	4.94E-07	5.88E+01	5.89E-06	1.98E-06	4.25E+01	3.40E-05
Rv2383c	mbtB	1.11E-17	2.08E+02	1.64E-15	1.22E-16	4.35E+01	2.32E-14
Rv2384	mbtA	5.76E-06	4.66E+01	4.95E-05	2.54E-05	33.6822168	0.000304834
Rv2397c	cysA1	0.005883294	36.75053885	0.016736668	0.001317277	55.08676826	0.008034986
Rv2400c	subI	3.32E-05	55.73715426	0.000211809	0.001423297	9.066718522	0.008589458
Rv2410c	-	9.64E-05	107.3199401	0.000525768	0.006156897	8.303452144	0.027999719
Rv2416c	eis	0.00015512	194.5424741	0.000788892	0.999998755	0.700824719	1
Rv2427c	proA	0.00013498	4.211520244	7.03E-04	8.46E-06	25.71639708	0.000118062
Rv2430c	PPE41	0.000181656	15.76025726	0.000899683	0.023110322	6.417210879	0.08020048
Rv2454c	-	0.001407168	44.08591967	0.004974495	0.001407168	44.08591967	0.008505025
Rv2455c	-	1.95E-06	6.83E+01	2.00E-05	3.74E-07	1.01E+02	8.09E-06
Rv2462c	tig	5.30E-09	1.24E+01	1.20E-07	9.07E-08	1.21E+01	2.30E-06
Rv2467	pepN	9.06E-07	1.56E+00	9.98E-06	1.13E-16	3.88E+01	2.26E-14
Rv2470	glbO	3.72E-06	1.25E+02	3.37E-05	3.72E-06	1.25E+02	5.83E-05
Rv2476c	gdh	4.12E-22	3.63E+00	1.26E-19	3.19E-17	1.35E+01	7.40E-15
Rv2492	-	2.44E-05	116.2848964	0.000163539	0.483014738	0.44463168	0.809502577
Rv2495c	pdhC	0.002768167	59.07618892	0.00884968	0.731088489	0.56335605	1
Rv2496c	pdhB	0.000166053	125.0740165	0.000833827	0.013391795	63.28213276	0.051758181
Rv2500c	fadE19	0.000814146	45.62356146	0.003196309	0.32642889	3.236926477	0.606072688
Rv2501c	accA1	2.25E-08	2.20E+00	4.33E-07	2.58E-09	1.54E+01	8.88E-08
Rv2506	-	0.38450528	0.753916334	0.567202336	0.00012973	116.2103828	0.001197285
Rv2510c	-	2.04E-05	96.77445435	0.000142244	0.071135768	0.810189342	0.188228177
Rv2518c	lppS	4.91E-07	3.97E+02	5.87E-06	4.91E-07	3.97E+02	1.00E-05
Rv2522c	-	0.002248932	23.74400689	0.007434749	0.205284926	1.08523725	0.434264975
Rv2527	-	0.00031439	19.26448057	0.001435508	0.004343404	10.14228403	0.021511478
Rv2535c	pepQ	2.29E-06	5.93E+01	2.25E-05	4.90E-07	7.74E+01	1.00E-05
Rv2536	-	0.002994807	89.6206161	0.009513546	0.009648862	11.27273048	0.039889319
Rv2543	lppA	3.81E-05	4.077832294	0.000236738	0.000392092	16.89926476	0.003093946
Rv2544	lppB	0.000100708	4.324556664	0.000545662	0.001576003	29.95406546	0.009354874
Rv2553c	-	1.88E-10	3.55E+02	6.40E-09	1.88E-10	3.55E+02	9.02E-09
Rv2563	-	0.000402631	189.1887383	0.001779181	0.000635668	32.13664039	0.004474425
Rv2564	glnQ	6.89E-06	5.31E+01	5.72E-05	5.86E-09	6.80E+01	1.83E-07
Rv2569c	-	2.24E-10	1.10E+02	7.37E-09	0.001670445	2.242957137	0.009842013
Rv2582	ppiB	0.00024831	66.84216208	0.001178437	0.00024831	66.84216208	0.002110108
Rv2583c	relA	8.45E-08	1.06E+02	1.38E-06	5.59E-08	1.15E+02	1.47E-06
Rv2605c	tesB2	0.000144106	104.3855757	0.000738165	0.000192526	11.45199159	0.001690232
Rv2606c	-	8.78E-06	1.81E+02	7.03E-05	2.32E-05	128.9410833	0.000282742
Rv2637	dedA	0.000849976	136.5166026	0.003307587	0.938504818	1.341691543	1

continued on next page

Rv2672	-	5.78E-05	1.980345954	3.40E-04	1.89E-06	2.46E+01	3.28E-05
Rv2680	-	0.000948521	10.9380681	0.003620219	0.000366669	103.9038752	0.002934092
Rv2681	-	0.001053167	1.455944396	3.93E-03	1.49E-05	26.19811066	0.000190699
Rv2683	-	3.25E-06	2.58E+01	2.99E-05	0.114266842	4.353543544	0.276254851
Rv2687c	-	0.0010805	15.96410799	0.003997346	0.101820749	1.083256913	0.251319586
Rv2690c	-	7.19E-05	153.853478	4.09E-04	7.19E-05	153.853478	0.000733271
Rv2694c	-	0.000836471	53.1940147	0.003264616	0.078958808	2.105291162	0.20443957
Rv2702	ppgK	0.001908351	29.05973879	0.006442708	0.110051674	0.871381192	0.267038137
Rv2710	sigB	1.53E-09	2.87E+02	3.97E-08	1.64E-10	3.13E+02	7.98E-09
Rv2716	-	0.000135223	18.65005722	0.000702983	0.101867531	2.214111322	0.251319586
Rv2717c	-	5.51E-07	1.66E+02	6.47E-06	0.000687806	17.48209287	0.004765516
Rv2733c	-	2.66E-07	3.46E+01	3.71E-06	1.48E-06	2.23E+01	2.65E-05
Rv2734	-	1.18E-05	1.98E+02	8.93E-05	0.002052041	21.05373826	0.011637129
Rv2735c	-	1.87E-11	1.05E+01	8.25E-10	0.000340281	1.350487288	0.002761828
Rv2737c	recA	2.91E-15	1.86E+00	2.57E-13	4.90E-14	1.33E+01	6.10E-12
Rv2744c	35kd_ag	0.001064552	314.8165875	0.003964162	0.063088603	8.156626091	0.170450662
Rv2752c	-	2.11E-06	1.80E+02	2.11E-05	0.000103818	20.22619622	0.000992513
Rv2776c	-	0.000992112	0.914734764	0.003746252	0.000431552	19.65840267	0.00333475
Rv2795c	-	5.24E-06	8.70E+01	4.54E-05	0.0015204	86.66781347	0.009065413
Rv2807	-	0.000367704	2.089468763	1.65E-03	7.35E-05	37.70196954	0.000745833
Rv2809	-	2.40E-05	365.4666194	0.000161231	0.022068194	5.327160798	0.077326175
Rv2816c	-	3.92E-05	13.23259535	0.000242061	0.154045554	1.013746889	0.347892771
Rv2817c	-	0.001046489	78.60073198	0.003915227	0.107346214	4.614164031	0.261751008
Rv2819c	-	1.88E-06	1.98E+02	1.94E-05	0.000482539	14.2117476	0.003635004
Rv2821c	-	4.63E-05	15.55839033	0.000280679	0.533874782	1.4757823	0.867302945
Rv2824c	-	3.21E-06	1.68E+00	2.96E-05	2.21E-06	1.55E+01	3.77E-05
Rv2833c	ugpB	2.74E-05	76.79957496	0.000180917	0.000550215	45.14733304	0.003999163
Rv2849c	cobO	0.002430807	166.4082249	0.007930532	0.52254259	3.295716105	0.854854744
Rv2855	mtr	4.80E-06	6.94E+01	4.20E-05	0.005598404	35.08984071	0.026218636
Rv2857c	-	0.000995672	236.0139052	0.003746252	0.164750958	80.43791309	0.368512126
Rv2869c	-	1.50E-09	1.30E+02	3.92E-08	2.99E-10	1.55E+02	1.35E-08
Rv2894c	xerC	0.001112296	104.509711	0.004084582	0.01210163	44.3128068	0.047604534
Rv2913c	-	2.07E-05	21.14189764	0.000143944	0.006663739	2.306878179	0.02964395
Rv2922c	smc	6.50E-06	6.94E+00	5.45E-05	1.03E-07	4.11E+01	2.56E-06
Rv2928	tesA	1.21E-08	3.46E+02	2.51E-07	0.011307615	6.382131787	0.045322308
Rv2930	fadD26	6.87E-19	1.25E+01	1.24E-16	1.36E-05	11.22146053	0.000175732
Rv2931	ppsA	1.38E-36	3.13E+01	5.47E-33	6.04E-12	8.08E+00	4.71E-10
Rv2932	ppsB	6.68E-28	7.29E+00	4.43E-25	5.81E-07	1.69E+01	1.16E-05
Rv2934	ppsD	2.64E-26	1.25E+01	1.50E-23	2.12E-07	1.58E+01	4.89E-06
Rv2935	ppsE	2.06E-35	5.47E+00	4.09E-32	7.02E-10	1.26E+01	2.97E-08
Rv2937	drrB	6.60E-05	1.325655943	3.80E-04	4.77E-07	1.60E+01	9.88E-06
Rv2939	papA5	1.89E-18	1.37E+01	3.01E-16	5.35E-05	5.677796332	0.000568593
Rv2940c	mas	1.16E-33	1.03E+01	1.15E-30	1.59E-09	7.40E+00	6.26E-08
Rv2945c	lppX	1.76E-06	4.12E+02	1.83E-05	4.48E-07	1.82E+01	9.38E-06
Rv2952	-	1.97E-06	4.96E+01	2.02E-05	0.011327728	3.351092984	0.045322308
Rv2959c	-	0.00016202	1.094570225	8.19E-04	3.62E-05	10.65495927	0.000402073
Rv2960c	-	0.00091943	94.50984117	0.003522711	0.541687263	1.016029283	0.872897461
Rv2967c	pca	1.52E-21	3.38E+00	4.31E-19	1.08E-23	4.98E+01	7.15E-21
Rv2985	mutT1	7.01E-06	2.23E+02	5.81E-05	0.001083608	3.600168485	0.006829652
Rv2997	-	0.001011731	19.34704331	0.003799483	0.032415443	1.319371908	0.10426132
Rv3005c	-	0.012393662	1.045580341	3.17E-02	3.32E-09	9.20E+01	1.09E-07
Rv3032	-	0.000731937	104.1026857	0.002911097	0.000731937	104.1026857	0.00498444
Rv3036c	TB22.2	2.21E-07	9.24E+01	3.17E-06	2.26E-09	5.10E+01	8.26E-08
Rv3040c	-	2.11E-05	95.91664279	1.46E-04	4.17E-05	104.6111858	0.000452697
Rv3041c	-	5.59E-05	100.0427651	0.000332162	0.000487377	20.3892083	0.003650276
Rv3044	fecB	2.24E-05	105.5304474	1.53E-04	2.24E-05	105.5304474	0.000273525
Rv3045	adhC	0.010656316	1.275776175	0.027992184	0.000926331	35.28302194	0.006031266
Rv3052c	nrdI	1.52E-05	466.7559545	1.11E-04	1.52E-05	466.7559545	0.000193754
Rv3057c	-	0.000131357	1.572678609	6.87E-04	2.16E-09	1.67E+01	8.02E-08
Rv3058c	-	0.004623356	1.361538521	0.013670695	0.00019715	13.03000512	0.001719442
Rv3066	-	0.000413539	22.26419293	0.001823333	0.005766116	2.530182898	0.026633966
Rv3071	-	1.09E-05	1.35E+02	8.43E-05	0.073096143	2.71012443	0.191756834
Rv3088	-	5.39E-05	18.7782012	0.000322328	0.065857416	1.215840582	0.176850062
Rv3101c	ftsX	2.75E-06	2.90E+02	2.63E-05	1.76E-05	130.4790612	0.000221635
Rv3102c	ftsE	0.007474086	53.63492472	0.020542116	0.000434348	148.1989031	0.00333475
Rv3106	fprA	0.000494913	43.09500864	0.002111876	0.695984522	1.06026645	1

continued on next page

Rv3127	-	4.89E-07	1.17E+02	5.87E-06	0.008455409	1.727070353	0.035659769
Rv3131	-	0.001528081	43.7335347	0.00530758	0.062304524	1.396625252	0.168676032
Rv3132c	devS	2.89E-07	4.33E+00	3.97E-06	3.73E-10	3.72E+01	1.65E-08
Rv3139	fadE24	5.39E-08	1.05E+01	9.16E-07	2.55E-06	2.01E+02	4.21E-05
Rv3147	nuoC	2.14E-05	58.89235247	0.000147635	0.022769482	0.514167083	0.079225048
Rv3148	nuoD	0.002274816	55.61913112	0.007501612	0.058002069	5.567300674	0.159636146
Rv3150	nuoF	4.42E-10	1.14E+02	1.38E-08	2.14E-05	1.077629699	0.00026389
Rv3151	nuoG	4.41E-07	9.33E+01	5.50E-06	0.00168258	7.287247818	0.009884228
Rv3159c	PPE53	2.16E-05	9.13463569	0.000148274	0.000358015	11.64500829	0.002876414
Rv3164c	moxR3	0.003112272	114.3291152	0.009807849	0.01172412	20.57470475	0.046667732
Rv3166c	-	0.000378705	94.58008328	0.001692256	0.009609552	5.339999304	0.039768146
Rv3168	-	4.52E-06	2.90E+02	4.00E-05	0.106809292	10.84224961	0.260761542
Rv3169	-	0.000982425	139.9665086	0.003723145	0.256733917	0.762961559	0.503510598
Rv3171c	hpx	0.000118377	61.70242576	0.000626877	0.105432411	1.029722219	0.258351631
Rv3176c	mesT	0.001618831	1.430524988	5.57E-03	3.96E-06	3.67E+01	6.13E-05
Rv3193c	-	1.20E-10	3.92E+01	4.42E-09	2.63E-09	4.33E+01	8.95E-08
Rv3194c	-	7.40E-07	2.80E+01	8.31E-06	1.20E-06	1.42E+01	2.23E-05
Rv3200c	-	0.000518034	35.52087646	2.18E-03	2.35E-07	5.56E+01	5.37E-06
Rv3205c	-	2.95E-05	40.90933953	1.92E-04	4.43E-05	39.23207714	0.000479056
Rv3207c	-	1.33E-05	3.85E+01	9.85E-05	0.033717464	8.686853745	0.107533564
Rv3209	-	0.003000168	14.42475364	0.009522481	0.083399049	3.533385449	0.213847853
Rv3211	rhlE	0.001915786	126.6007432	0.006454583	0.006465074	88.24784331	0.029052655
Rv3214	gpm2	0.000119167	53.57784143	0.000630222	0.002963123	10.15363494	0.015754466
Rv3223c	sigH	4.63E-07	3.24E+01	5.70E-06	3.22E-09	1.80E+01	1.07E-07
Rv3245c	mtrB	0.000265302	26.32329421	0.0012401	0.000265302	26.32329421	0.002216611
Rv3249c	-	0.003160479	10.77502262	0.009936147	0.000517166	18.983036	0.0038159
Rv3261	fbiA	0.000964567	89.25065377	0.003667381	0.0003954	101.5755439	0.003113871
Rv3262	fbiB	0.001219254	109.5847384	0.004420211	0.003050233	79.16047518	0.016152831
Rv3267	-	8.66E-14	3.41E+02	5.74E-12	6.80E-12	1.41E+02	5.20E-10
Rv3268	-	0.00019118	9.377689682	9.40E-04	7.60E-07	4.81E+01	1.47E-05
Rv3270	ctpC	0.000164216	13.50376523	8.26E-04	1.59E-06	3.05E+01	2.83E-05
Rv3274c	fadE25	5.57E-05	125.4309305	3.31E-04	8.65E-06	232.3153528	0.00011904
Rv3282	maf	0.000266721	43.05470517	0.001245012	0.005644003	37.76013209	0.026252864
Rv3283	sseA	2.53E-09	2.58E+01	6.21E-08	7.61E-13	4.63E+02	7.39E-11
Rv3286c	sigF	0.002446153	95.7398064	0.007974059	0.516724164	1.766843251	0.846729295
Rv3291c	-	0.002380725	258.6871318	0.00779271	0.024261571	24.295559	0.083107896
Rv3300c	-	0.000163084	56.08841457	0.000822032	0.145208502	1.686069653	0.333041645
Rv3303c	lpdA	7.93E-06	6.64E+01	6.46E-05	1.14E-05	23.39987451	0.000151192
Rv3305c	amiA1	3.52E-08	1.37E+02	6.39E-07	0.000462518	8.694042842	0.003510371
Rv3311	-	1.25E-08	3.44E+02	2.57E-07	5.78E-09	3.03E+02	1.83E-07
Rv3316	sdhC	0.001948772	50.76858661	0.006545832	0.255919556	5.605128353	0.503405959
Rv3323c	moaX	2.12E-06	4.67E+02	2.12E-05	0.000342174	3.433947654	0.002771542
Rv3340	metC	4.20E-07	1.93E+01	5.31E-06	4.13E-09	2.22E+01	1.35E-07
Rv3342	-	0.000466327	91.50998659	0.00201585	0.022288906	10.75716185	0.077812434
Rv3371	-	0.001222395	11.68742276	0.004427562	0.563360461	2.734849526	0.900516299
Rv3378c	-	4.34E-06	7.72E+01	3.84E-05	0.003411779	7.985855222	0.017713633
Rv3383c	idsB	1.33E-07	1.97E+01	2.02E-06	2.23E-05	4.741222584	0.000273525
Rv3389c	-	0.000394054	143.8777636	0.001751009	0.043120775	11.038092	0.127625076
Rv3394c	-	0.047344306	1.844784901	0.097306617	0.000492259	11.46210777	0.003673006
Rv3395A	-	0.001190255	96.87619585	0.004327356	0.198050872	9.77961002	0.426216622
Rv3399	-	1.85E-07	3.98E+01	2.71E-06	4.08E-06	1.61E+02	6.21E-05
Rv3400	-	0.00034615	5.255245488	1.57E-03	5.81E-06	7.58E+01	8.46E-05
Rv3412	-	0.000675872	154.1777543	0.002728875	0.179941582	3.465949569	0.397129673
Rv3414c	sigD	0.003097682	43.93870291	0.009769614	1	0.259960432	1
Rv3417c	groEL	0.000476466	35.75800251	0.002051482	0.004746461	15.83805486	0.023133181
Rv3420c	rimI	1.43E-08	1.14E+02	2.93E-07	0.00051391	10.49677006	0.003798924
Rv3421c	-	0.000221157	42.71015893	0.001062249	0.441949456	1.717767803	0.756947883
Rv3424c	-	0.002340083	113.1647006	0.007690908	0.251321859	0.927175901	0.49726718
Rv3427c	-	0.276760765	0.52217404	0.431299985	0.00096147	159.5440345	0.006207411
Rv3434c	-	4.59E-05	46.11263872	0.000279194	0.01303127	3.434387318	0.050681471
Rv3448	-	2.02E-06	1.41E+01	2.04E-05	0.000193333	8.161778236	0.001693582
Rv3454	-	4.11E-07	7.07E+01	5.23E-06	0.008389954	1.55829782	0.035439433
Rv3484	cpsA	1.52E-06	1.64E+00	1.59E-05	4.90E-17	2.56E+01	1.02E-14
Rv3491	-	6.21E-05	182.2869975	0.000359964	0.272601346	4.885116493	0.527547907
Rv3493c	-	9.64E-05	129.5365156	0.000525768	0.006866313	9.64870487	0.030307797
Rv3494c	mce4F	1.02E-09	4.70E+00	2.91E-08	1.80E-09	1.25E+01	6.93E-08

continued on next page

Rv3496c	mce4D	0.000800149	1.329488087	3.14E-03	2.51E-08	1.01E+01	7.04E-07
Rv3497c	mce4C	0.000499899	17.62665716	0.002127061	0.002819488	10.84683026	0.015193908
Rv3498c	mce4B	0.001537049	15.43114231	0.005334069	0.000919112	16.90615985	0.006021927
Rv3502c	fabG	4.79E-07	1.62E+02	5.82E-06	4.41E-06	1.12E+01	6.62E-05
Rv3527	-	0.038958209	5.821724167	0.08361403	0.00088387	55.77183385	0.005810169
Rv3528c	-	0.000208641	23.70869819	0.001009449	0.196452925	5.545586729	0.424385271
Rv3533c	PPE62	3.72E-07	1.45E+02	4.87E-06	0.003475865	8.24275282	0.017975963
Rv3536c	-	0.000640259	18.01378143	2.59E-03	2.42E-05	93.49752079	0.000292022
Rv3537	-	3.50E-14	1.82E+01	2.53E-12	2.55E-09	1.38E+01	8.88E-08
Rv3540c	ltp2	0.001407168	71.11777411	0.004974495	0.059857294	9.768580122	0.163609938
Rv3542c	-	0.001772217	17.42590792	0.006029178	0.001009201	16.15102191	0.006452723
Rv3543c	fadE29	8.42E-07	4.26E+01	9.30E-06	5.85E-07	4.47E+01	1.16E-05
Rv3544c	fadE28	0.000260002	3.672925821	1.23E-03	3.60E-05	14.79066979	0.000401194
Rv3548c	-	0.002394611	127.8964093	0.007831718	0.005282557	126.5363078	0.025070084
Rv3552	-	0.000731937	38.88020387	0.002911097	0.015834689	16.94083926	0.059242294
Rv3553	-	0.003097682	130.111872	0.009769614	0.296907295	10.86284154	0.562553746
Rv3555c	-	1.17E-05	1.00E+02	8.93E-05	0.00020805	9.950857572	0.001790941
Rv3556c	fadA6	8.07E-06	5.30E+01	6.56E-05	2.30E-06	5.77E+01	3.88E-05
Rv3559c	-	0.008249301	53.76854354	0.022335877	0.00045494	58.40007005	0.003459461
Rv3560c	fadE30	5.46E-05	3.046621132	3.26E-04	8.52E-06	12.12240323	0.000118265
Rv3561	fadD3	3.71E-05	110.2355065	0.000232508	0.429457315	1.089687388	0.740334521
Rv3567c	-	0.000585995	150.4029208	0.002410032	0.120791407	4.975667159	0.288174819
Rv3568c	bphC	0.001871028	70.0843516	0.006332832	0.011304803	14.83751415	0.045322308
Rv3569c	bphD	0.003966989	24.72380187	0.012034107	0.000649374	49.17512635	0.004554778
Rv3571	hmp	8.97E-05	85.80885459	0.000496666	0.07203618	7.169061242	0.189629793
Rv3574	-	0.000330613	68.16275892	0.001504403	0.006723167	22.29530554	0.029834706
Rv3577	-	4.60E-05	121.7176386	0.000279349	0.013950396	23.08545034	0.053516552
Rv3578	arsB2	0.002842032	1.108979699	0.009071236	0.000454309	45.31489212	0.003459461
Rv3591c	-	0.059233811	1.615894174	0.117025765	0.001498086	11.61795761	0.00894578
Rv3610c	ftsH	1.97E-05	0.739189498	1.38E-04	2.30E-06	1.48E+01	3.88E-05
Rv3615c	-	0.000815471	8.202016696	3.20E-03	1.88E-05	12.7098927	0.0002345
Rv3616c	-	3.19E-07	6.01E+00	4.31E-06	4.16E-07	2.23E+01	8.90E-06
Rv3626c	-	0.000524315	24.60450357	0.002204231	0.018733647	6.337345032	0.068039921
Rv3645	-	1.12E-10	3.35E+02	4.15E-09	1.12E-10	3.35E+02	5.62E-09
Rv3651	-	0.001648612	1.874877145	5.66E-03	5.69E-05	16.66249563	0.000599136
Rv3664c	dppC	9.85E-05	40.93903836	0.000535684	0.103016082	3.47892679	0.253681088
Rv3665c	dppB	3.36E-08	8.74E+00	6.15E-07	1.81E-09	2.70E+02	6.93E-08
Rv3666c	dppA	4.21E-07	1.99E+02	5.31E-06	2.04E-05	23.122938	0.000251649
Rv3668c	-	0.000113621	40.58451058	0.000604914	0.000709067	11.45713153	0.004870395
Rv3670	ephE	0.000122158	154.1461221	6.44E-04	2.60E-05	228.5547665	0.000310462
Rv3671c	-	0.003781909	84.75387835	0.01151658	0.001317277	116.8024814	0.008034986
Rv3679	-	0.000910059	5.11165741	3.49E-03	1.27E-06	4.01E+02	2.33E-05
Rv3680	-	3.07E-07	2.65E+00	4.19E-06	2.68E-12	4.35E+02	2.27E-10
Rv3682	ponA2	6.36E-19	6.84E+01	1.20E-16	1.20E-25	6.68E+02	1.20E-22
Rv3683	-	0.000200165	2.83888168	9.72E-04	3.30E-05	17.68668158	0.000374744
Rv3684	-	8.92E-08	2.40E+02	1.44E-06	0.000816976	2.396663203	0.005401166
Rv3699	-	0.000546692	16.1447759	0.002283819	0.008283798	2.106388575	0.035047517
Rv3710	leuA	0.001064465	1.591731383	3.96E-03	1.00E-05	57.75740299	0.000135543
Rv3717	-	0.001582957	3.050370227	5.47E-03	8.38E-06	152.1326452	0.000117392
Rv3720	-	3.00E-08	1.19E+00	5.66E-07	8.98E-08	1.20E+01	2.29E-06
Rv3757c	proW	0.001190377	16.53315409	0.004327356	0.296314784	0.894875141	0.562047964
Rv3759c	proX	0.000384675	79.36188359	0.00171316	0.130038969	1.677462031	0.304752492
Rv3777	-	0.00041681	1.271752856	0.001829641	0.000490821	118.8285309	0.00366916
Rv3802c	-	6.05E-11	3.72E+02	2.38E-09	6.05E-11	3.72E+02	3.44E-09
Rv3804c	fbpA	3.56E-06	9.04E+01	3.23E-05	0.001174391	3.580144032	0.007288415
Rv3810	pirG	0.000114168	1.606948628	6.07E-04	3.80E-08	1.96E+02	1.03E-06
Rv3813c	-	0.001172567	84.22665859	0.004274333	0.035998736	80.20532277	0.112552653
Rv3823c	mmpL8	5.43E-24	3.14E+00	1.80E-21	4.38E-27	3.10E+01	8.71E-24
Rv3838c	pheA	0.004999953	17.13323881	0.014556964	0.000995672	25.55697785	0.006406283
Rv3860	-	0.00025343	2.280031044	1.20E-03	8.28E-06	11.35856831	0.00011677
Rv3864	-	2.36E-07	1.82E+02	3.36E-06	2.69E-08	2.73E+02	7.49E-07
Rv3868	-	2.63E-15	3.69E+01	2.44E-13	5.81E-14	1.77E+01	7.00E-12
Rv3870	-	2.23E-12	4.61E+00	1.20E-10	1.45E-12	1.46E+01	1.28E-10
Rv3876	-	4.70E-08	1.45E+01	8.23E-07	4.11E-06	1.74E+00	6.21E-05
Rv3916c	-	3.07E-05	51.37038448	1.98E-04	3.72E-06	7.48E+01	5.83E-05
Rv3919c	gidB	0.001230753	2.316777795	4.45E-03	4.09E-05	32.20429099	0.000446352

Table 3.2. Genes required for surviving CD4-mediated stress

Rv Number	Gene	MWU P-value	wt:MHC ratio
Rv0019c	-	0.036291718	0.021305927
Rv0050	ponA1	0.039678686	0.112297814
Rv0163	-	0.008756427	0.210237414
Rv0176	-	0.007152225	0.103444727
Rv0200	-	0.00535539	0.056981315
Rv0211	pckA	0.008215453	0.25
Rv0241c	-	0.007548983	0.25
Rv0242c	fabG	0.001066115	0.25
Rv0634A	-	0.003395271	0
Rv0665	-	0.044644399	0
Rv0784	-	0.011561192	0
Rv0893c	-	0.036165907	0.25
Rv0926c	-	0.046163388	0.108115878
Rv0937c	-	0.014453174	0.199901486
Rv1062	-	0.025530204	0.062573889
Rv1096	-	0.012651451	0.100469202
Rv1121	zwf1	0.035478742	0.109182797
Rv1170	mshB	0.006312207	0
Rv1193	fadD36	0.026756502	0.054452155
Rv1237	sugB	0.011038856	0
Rv1432	-	0.045176818	0.171597184
Rv1448c	tal	0.012186994	0.012872952
Rv1512	epiA	0.003548422	0
Rv1519	-	0.015260332	0
Rv1568	bioA	0.007400462	0.190061243
Rv1604	impA	0.021724434	0.25
Rv1609	trpE	0.001375862	0.207867048
Rv1622c	cydB	5.53E-06	0.200941859
Rv1755c	plcD	0.003334523	0
Rv1829	-	0.004701583	0.146164946
Rv1913	-	0.012296325	0.080949093
Rv1954c	-	0.01299173	0.244898291
Rv1962c	-	0.010629303	0.208252337
Rv2135c	-	0.027861141	0
Rv2192c	trpD	0.037669967	0.25
Rv2222c	glnA2	0.00430293	0.023746514
Rv2269c	-	0.016223977	0.009559459
Rv2274c	-	0.001735063	0
Rv2383c	mbtB	0.001084174	0.059279337
Rv2384	mbtA	0.017657347	0
Rv2496c	pdhB	0.018302614	0.023560933
Rv2553c	-	0.001520846	0.25
Rv2606c	-	0.019274815	0
Rv2745c	-	0.032393036	0.185798859
Rv2809	-	0.003259852	0
Rv2857c	-	0.011038856	0
Rv2879c	-	0.01002359	0
Rv2931	ppsA	0.003091397	0.199297953
Rv2966c	-	0.040451605	0.111901222
Rv3101c	ftsX	0.000202429	0
Rv3168	-	0.01784547	0
Rv3267	-	0.000888031	0.25
Rv3271c	-	0.022389724	0
Rv3320c	-	0.011004546	0.20725692
Rv3502c	fabG	0.011930481	0.17299943
Rv3636	-	0.005596797	0.061449799
Rv3684	-	0.004093713	0.209292629
Rv3868	-	0.002439503	0.241047928

Table 3.3A. Genes required for surviving acid stress (Tyloxapol pH 6.5 v Tyloxapol pH 4.5)

Rv Number	Gene	MWU P-value	Ty6.5:Ty4.5 ratio
Rv1621c	cydD	0.011403588	21.83314743
Rv1623c	cydA	0.015494229	20.62578152
Rv1622c	cydB	0.018733271	16.53171566
Rv1620c	cydC	0.000275637	15.46023648

Table 3.3B. Genes required for surviving acid stress in phosphate-citrate buffer (Tyloxapol pH 6.5 v pcit pH 4.5)

Rv	gene	MWU P-value	Ty6.5:pcit4.5 ratio
Rv0536	galE3	2.01E-06	21.10052984
Rv2282c	-	5.42E-05	20.93847196
Rv1339	-	0.028430726	20.57718069
Rv2665	-	0.016300213	19.37981555
Rv2943A	-	0.007903478	17.49904946
Rv1717	-	0.002628449	17.47602033
Rv1620c	cydC	0.000121797	17.31160396
Rv0326	-	0.047650855	17.15444784
Rv1621c	cydD	0.025758844	16.77289417
Rv1623c	cydA	0.011462012	15.69509581
Rv2630	-	0.017523493	14.87833358
Rv0612	-	0.001483126	14.55994587
Rv1287	-	0.000127653	13.96261956
Rv2758c	-	0.022042404	13.00406216
Rv3229c	-	0.000104834	12.90834055
Rv3203	lipV	0.017364817	11.79062897
Rv2089c	pepE	0.01883695	11.69807278
Rv3013	-	0.001211151	11.62492734
Rv1622c	cydB	0.002570576	11.20090151
Rv1284	-	0.034114539	10.73117748
Rv0324	-	0.001112296	10.08128541

Table 3.3C. Genes required for surviving nitrosative stress

Rv Number	Gene	MWU P-value	pH7:DETANO ratio
Rv0757	phoP	2.91E-05	23.03266802
Rv3283	sseA	2.95E-05	13.57730202
Rv3270	ctpC	1.24E-06	12.85968522
Rv1620c	cydC	0.000126385	12.17374061
Rv1622c	cydB	8.38E-05	9.333274474
Rv2563	-	0.000142698	8.034889817
Rv0467	icl	3.31E-05	7.78349238
Rv3855	ethR	7.49E-05	7.339175141
Rv0561c	-	9.97E-05	6.851134457
Rv3200c	-	1.19E-05	6.747794845
Rv2476c	gdh	1.69E-18	6.335121997
Rv2047c	-	3.95E-06	6.046666345

Table 3.3D. Genes required for surviving tryptophan starvation

Rv Number	gene	MWU P-value	trp:Rv ratio
Rv2192c	trpD	0.01536	258
Rv2246	kasB	0.04024	193
Rv1612	trpB	0.04017	142
Rv3160c	-	0.01624	101
Rv1609	trpE	0.00435	64.167
Rv1053c	-	0.04468	23.227
Rv3374	echA18.1	0.03826	23.111
Rv2661c	-	0.04043	18.444
Rv1559	ilvA	0.01168	18.148
Rv1013	pks16	0.00002	17.809
Rv2283	-	0.04043	15.889
Rv0346c	ansP2	0.00096	11.155

Section 3.4: Bibliography

1. WHO. (http://www.who.int/tb/publications/global_report/2011/en/, 2011).

2. Ernst, J.D. The immunological life cycle of tuberculosis. *Nat Rev Immunol* **12**, 581-91 (2012).

3. Ehrt, S. & Schnappinger, D. Mycobacterial survival strategies in the phagosome: defence against host stresses. *Cell Microbiol* **11**, 1170-8 (2009).

4. Flynn, J.L. Lessons from experimental Mycobacterium tuberculosis infections. *Microbes Infect* **8**, 1179-88 (2006).

5. Cosgrove, D. et al. Mice Lacking MHC Class II Molecules. *Cell* **66**, 1051-1066 (1991).

6. Grusby, M.J., Johnson, R.S., Papaioannou, V.E. & Glimcher, L.H. Depletion of CD4+ T Cells in Major Histocompatibility Complex Class II-Deficient Mice. *Science* **253**, 1417-1420 (1991).

7. Caruso, A.M. et al. Mice Deficient in CD4 T Cells Have Only Transiently Diminished Levels of IFN-{gamma}, Yet Succumb to Tuberculosis. *Journal of immunology (Baltimore, Md : 1950)* **162**, 5407-5416 (1999).

8. Scanga, C.A. et al. Depletion of CD4+ T Cells Causes Reactivation of Murine Persistent Tuberculosis Despite Continued Expression of Interferon-gamma and Nitric Oxide Synthase 2. *Journal of Experimental Medicine* **192**, 347-358 (2000).

9. Mogues, T., Goodrich, M.E., Ryan, L., Lacourse, R. & North, R.J. The Relative Importance of T Cell Subsets in Immunity and Immunopathology of Airborne Mycobacterium tuberculosis Infection in Mice. *Journal of Experimental Medicine* **193**, 271-280 (2001).

10. McDermid, J.M. et al. Host iron redistribution as a risk factor for incident tuberculosis in HIV infection: an 11-year retrospective cohort study. *BMC Infect Dis* **13**, 48 (2013).

11. Macpherson, P. et al. Risk factors for mortality in smear-negative tuberculosis suspects: a cohort study in Harare, Zimbabwe. *Int J Tuberc Lung Dis* **15**, 1390-6 (2011).

12. Pawlowski, A., Jansson, M., Skold, M., Rottenberg, M.E. & Kallenius, G. Tuberculosis and HIV co-infection. *PLoS Pathog* **8**, e1002464 (2012).

13. Selwyn, P.A. et al. A Prospective Study of the Risk of Tuberculosis Among Intravenous Drug Users with Human Immunodeficiency Virus Infection. *The New England Journal of Medicine* **320**, 545-550 (1989).

14. Flynn, J.L. et al. Tumor Necrosis Factor-alpha Is Required in the Protective Immune Response Against Mycobacteriumtuberculosis in Mice. *Immunity* **2**, 561-572 (1995).

15. Bold, T.A., Banaei, N., Wolf, A.J. & Ernst, J.D. Suboptimal Activation of Antigen-Specific CD4 Effector Cells Enables Presistence of *M. tuberculosis* In Vivo. *PLoS Pathogens* **7**, e1002063 (2011).

16. Flynn, J.L. et al. An Essential Role for Interferon gamma in Resistance to Mycobacterium tuberculosis Infection. *Journal of Expermental Medicine* **178** (1993).

17. Gallegos, A.M. et al. A Gamma Interferon Independent Mechanism of CD4 T Cell Mediated Control of M. tuberculosis Infection in vivo. *PLoS Pathogens* **7**, e1002052. (2011).

18. Russell, D.G. Who puts the tubercle in tuberculosis? *Nat Rev Microbiol* **5**, 39-47 (2007).

19. Cooper, A.M. et al. Disseminated Tuberculosis in Interferon-gamma Gene-disrupted Mice. *Journal of Experimental Medicine* **178**, 2243-2247 (1993).

20. Zhang, Y.J. et al. Global assessment of genomic regions required for growth in Mycobacterium tuberculosis. *PLoS Pathog* **8**, e1002946 (2012).

21. Sassetti, C.M. & Rubin, E.J. Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**, 12989-94 (2003).

22. Darwin, K.H. & Nathan, C.F. Role for nucleotide excision repair in virulence of Mycobacterium tuberculosis. *Infect Immun* **73**, 4581-7 (2005).

23. McKinney, J.D. et al. Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**, 735-738 (2000).

24. Munoz-Elias, E.J. & McKinney, J.D. Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nat Med* **11**, 638-44 (2005).

25. Marrero, J., Rhee, K.Y., Schnappinger, D., Pethe, K. & Ehrt, S. Gluconeogenic carbon flow of tricarboxylic acid cycle intermediates is

critical for Mycobacterium tuberculosis to establish and maintain infection. *Proc Natl Acad Sci U S A* **107**, 9819-24 (2010).

26. Shi, L. et al. Changes in energy metabolism of Mycobacterium tuberculosis in mouse lung and under in vitro conditions affecting aerobic respiration. *Proc Natl Acad Sci U S A* **102**, 15629-34 (2005).

27. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).

28. Griffin, J.E. et al. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* **7**, e1002251 (2011).

29. Goodman, A.L. et al. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**, 279-89 (2009).

30. Zomer, A., Burghout, P., Bootsma, H.J., Hermans, P.W. & van Hijum, S.A. ESSENTIALS: software for rapid analysis of high throughput transposon insertion sequencing data. *PloS one* **7**, e43012 (2012).

31. Barquist, L. et al. A comparison of dense transposon insertion libraries in the Salmonella serovars Typhi and Typhimurium. *Nucleic Acids Res* (2013).

32. Clatworthy, A.E., Pierson, E. & Hung, D.T. Targeting virulence: a new paradigm for antimicrobial therapy. *Nat Chem Biol* **3**, 541-8 (2007).

33. Cegelski, L., Marshall, G.R., Eldridge, G.R. & Hultgren, S.J. The biology and future prospects of antivirulence therapies. *Nat Rev Microbiol* **6**, 17-27 (2008).

Chapter 4.

Host Mechanisms of Tryptophan Starvation and Inhibition of Bacterial Tryptophan Biosynthesis to treat TB.

Section 4.1: Overview and Attributions

Overview. The results in Chapter 3 suggested that CD4 T cells attempted to kill Mtb by starving the pathogen of tryptophan, and that Mtb survival in the face of CD4 T cells was dependent on tryptophan biosynthesis. Thus, we sought 1.) to understand the mechanism by which CD4 T cells starve Mtb of tryptophan and 2.) to chemically inhibit bacterial tryptophan biosynthesis in order to help the immune system kill Mtb. We found that IFN-γ, a cytokine secreted by CD4 T cells during infection, was able to induce killing of tryptophan auxotrophs to a greater extent than wildtype Mtb. The auxotroph's hypersusceptibility to IFN-γ was dependent on IFN-γ stimulation of indoleamine-2,3-dioxygenase (IDO), which catabolizes intracellular tryptophan. CD4 T cells, through IDO induction by IFN-γ, starve Mtb of exogenous tryptophan, and we predicted that inhibiting endogenous tryptophan biosynthesis in this environment would be bactericidal. We found two halogenated anthranilates that could inhibit tryptophan biosynthesis, and showed that these compounds are potent inhibitors of bacterial growth, especially in synergy with immune activation.

Attributions. The following Chapter is an expanded version a manuscript, which also includes most of Chapter 3. I wrote the manuscript and we aim to submit in the next month. The work in this Chapter could not have been done without the generous help of Sam Behar and his graduate student Alissa Rothchild. Sam invited me into his lab, and Alissa and I did all of the mouse macrophage experiments together. We "discovered" the anthranilate compounds in a paper

from Laurie Rahme's group at the Massachusetts General Hospital, and she was kind enough to share her thoughts and valuable input. Joseph Mire, a graduate student in Jim Sacchettini's lab, and I were collaborating on another tryptophan-related project, and his expertise was tremendously helpful. Finally, Deeann Wallis in Jim Sacchettini's lab and Veronique Dartois at the Public Health Research Institute did the pharmacokinetic profiling in mice, and their help in designing the mouse efficacy study was invaluable. Brian Schuster, a technician in our lab, helped with delivering the compounds during the mouse efficacy study, and Andrej Trauner, a post-doc in the lab, tested compounds in *M. smegmatis*.

Section 4.2: Bacterial tryptophan biosynthesis determines Mtb survival upon IFN-γ induction of IDO

Introduction

The transposon screen and follow-up studies demonstrated that tryptophan biosynthesis is required for surviving CD4-mediated stress. This suggests that CD4 T cells are able to starve Mtb of tryptophan. Some pathogens, such as *Toxoplasma gondii* and certain strains of *Chlamydia*, are natural tryptophan auxotrophs, and we realized that tryptophan starvation during infection with these pathogens could be the same mechanisms that fail to kill the non-tryptophan auxotrophic Mtb[1-8]. In a cell culture model of *Chlamydia* infection, bacterial growth is inhibited by the addition of IFN-γ, and this growth inhibition can be almost entirely reversed by the addition of tryptophan[1]. Furthermore, chemical inhibition or genetic perturbation of IDO could also reverse this growth inhibition[2-4, 7]. IDO is one of the most highly transcriptionally induced genes with IFN-γ stimulation, and is required for the tryptophan starvation seen in IFN-γ-activated macrophages[5, 7]. We were interested in whether or not this same pathway was induced by CD4 T cells, and if IDO-mediated tryptophan starvation was responsible for the specific requirement for bacterial tryptophan biosynthesis in the face of CD4 immunity.

Results

Tryptophan auxotrophy is bactericidal

To study the effects of tryptophan auxotrophy and starvation during Mtb infection, we deleted the *trpE* gene in Mtb and replaced it with the hygromycin resistance gene (Fig 4.1A). TrpE, or Anthranilate synthase, converts chorismate to anthranilate in the first committed step of tryptophan biosynthesis. The next enzyme in the pathway, TrpD, converts anthranilate to *N*-(5'-phosphoribosyl)-anthranilate, and a knockout of *trpD* has been shown to be a tryptophan auxotroph as well (Fig 4.1B)[9, 10]. The *trpE* knockout grew normally in liquid media only when supplemented with 1 mM tryptophan (Fig 4.2A). Interestingly, lower amounts of tryptophan delayed entry into the logarithmic phase of growth, but the growth dynamics from that point onwards was similar to wildtype. Genetic complementation with the *trpE* gene restored normal growth in media lacking tryptophan. To confirm our screen results (Fig 3.2A), we infected wt and MHCII$^{-/-}$ mice with a mixture of wt Mtb, the *trpE* knockout and the complemented strain. Growth of these strains in mice are pending, with results expected in the coming weeks.

Auxotrophy in Mtb is not always bactericidal[9]. Our results suggested that CD4 cells help starve Mtb of exogenous tryptophan, so we sought to test the bactericidal potential of blocking endogenous tryptophan biosynthesis in the face of this exogenous starvation. To do so, we grew the auxotroph in tryptophan to both mid-log and stationary phase, and then continued the culture either in the presence or absence of tryptophan and plated to measure survival. The auxotroph was rapidly killed when starved of tryptophan, suggesting that

tryptophan biosynthesis would be a good bactericidal drug target (Fig 4.2B and C). Interestingly, the rate of death of this auxotroph was much more rapid compared to the *trpD* knockout in previous studies. In one paper, no loss of viable bacteria was measured at 10 days after tryptophan starvation of the *trpD* knockout, and a 100-fold effect was seen 30 days after starvation[9]. The *trpE* knockout, on the other hand, has about a 100,000-fold loss of viability at two weeks, a level which is not even reached at 13 weeks of *trpD* knockout starvation. The key difference between the two mutants is the ability of the *trpD* knockout to make anthranilate, and it is possible that there is an alternative pathway for tryptophan biosynthesis from anthranilate. Regardless, it appears that blocking tryptophan biosynthesis is much more bactericidal than previously appreciated.

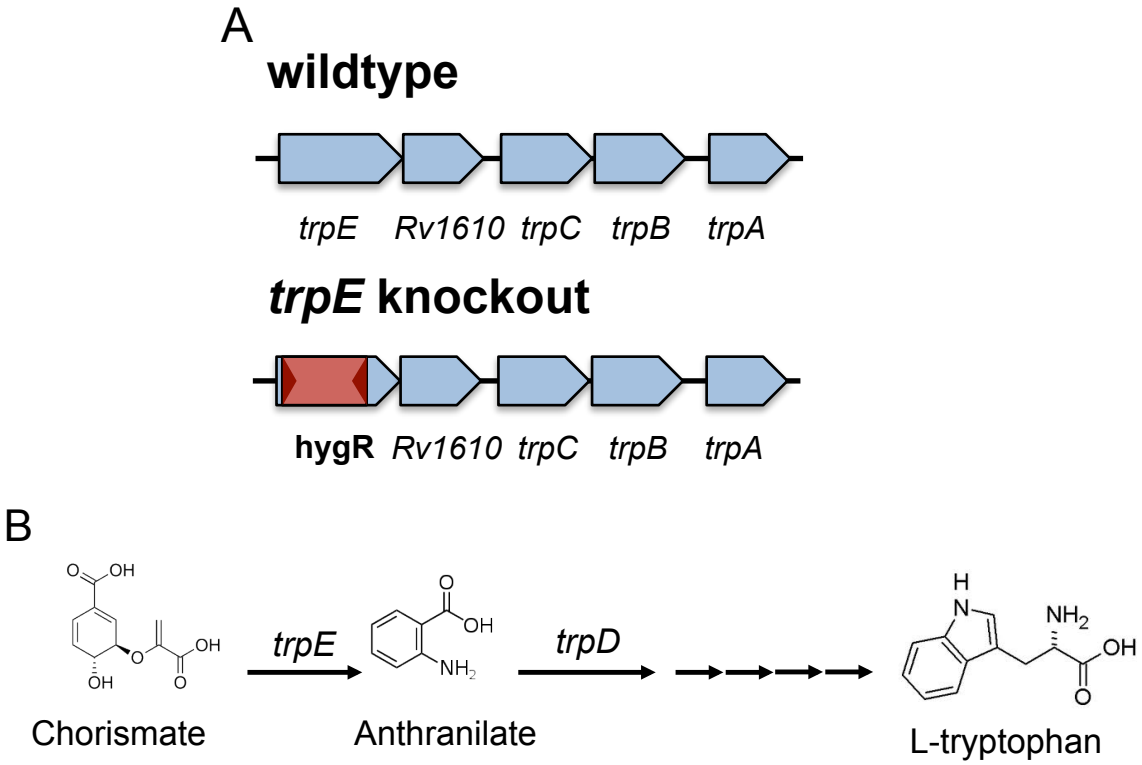


Figure 4.1. A.) The hygromycin resistance gene, flanked by 500 bp regions surrounding (and partially including) *trpE*, transformed into dsDNA recombinase-expressing Mtb, resulting in the construction of a *trpE* knockout strain. Transformants were plated on 1 mM tryptophan. B.) TrpE conversion of chorismate to anthranilate forms the first dedicated synthetic step for making tryptophan.

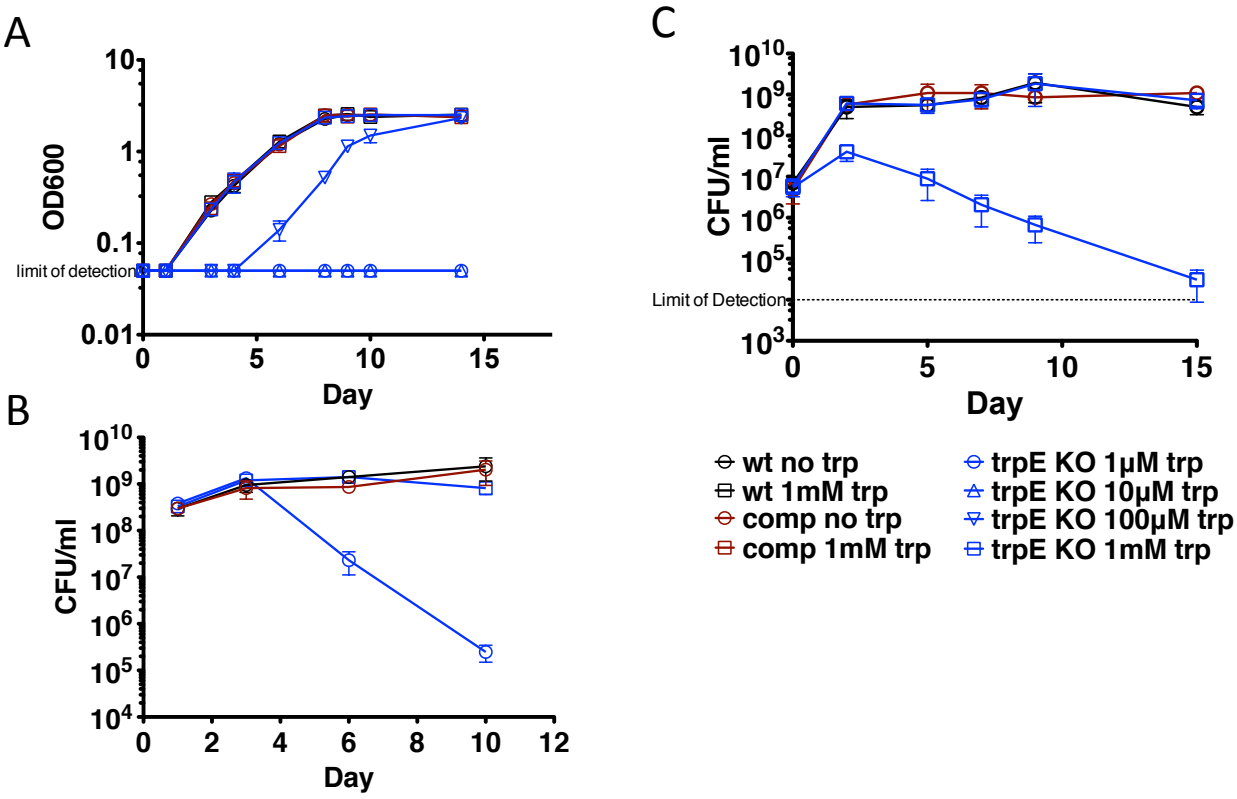


Figure 4.2. A.) Growth of the auxotroph (trpE KO), H37Rv wildtype (wt) and complemented strain (comp) in 7H9 media in the presence or absence of tryptophan was measured by OD600. To measure CFU/ml for trpE KO, wt, and comp strains grown to mid-log phase (B) or stationary phase (C), washed, and suspended in 7H9 media with or without tryptophan.

Tryptophan auxotrophs are hypersusceptible to IFN-γ mediated killing in macrophages

During a typical lung infection, Mtb first enters alveolar macrophages, which then form an inflammatory structure known as the granuloma[11]. Upon adaptive immune activation, CD4 T cells enter the granuloma and stimulate the infected macrophages[12]. Our screen results suggested that one of these stimulatory mechanisms demands the need for bacterial tryptophan biosynthesis, and we sought to delineate the mechanism by which CD4 cells exert this need. To do so, we infected cultured mouse peritoneal macrophages with wt Mtb, the *trpE* knockout and the complemented strain. Over a five-day infection, the wt and complemented increased in colony forming units (CFU) by about 5-fold, while the *trpE* knockout had no measureable growth, demonstrating that tryptophan biosynthesis is required for growth even in unstimulated macrophages (Fig 4.3A).

In addition to the growth defect in macrophages, we also expected that the *trpE* knockout would be especially sensitive to CD4-mediated stress. Indeed, the *trpE* knockout was significantly hypersusceptible to IFN-γ and TNF-α, cytokines whose secretion is increased upon the arrival of CD4 T cells to an Mtb lesion. Compared to unstimulated macrophages, stimulated macrophages decrease growth of wt Mtb by about 40%, where as IFN-γ and TNF-α stimulation decreased viable *trpE* knockout bacteria by about 90% (Fig 4.3B). We also simulated the effect of CD4 cells by co-culturing infected macrophages with CD4 cells from spleens of Mtb-infected mice. The *trpE* knockout was also

hypersusceptible to the growth-inhibition of CD4 T cells in co-culture (Fig 4.4B). Since human macrophages differ from mouse macrophages in their mycobacterial killing strategies, we tested the growth of our Mtb strains in monocyte-derived macrophages from human donors. As with previous reports, we observed that IFN-γ stimulation alone does not measurably inhibit wt Mtb growth. However, while the *trpE* knockout grows slightly over the 5-day infection, its growth is inhibited by IFN-γ stimulation of macrophages (Fig 4.5).

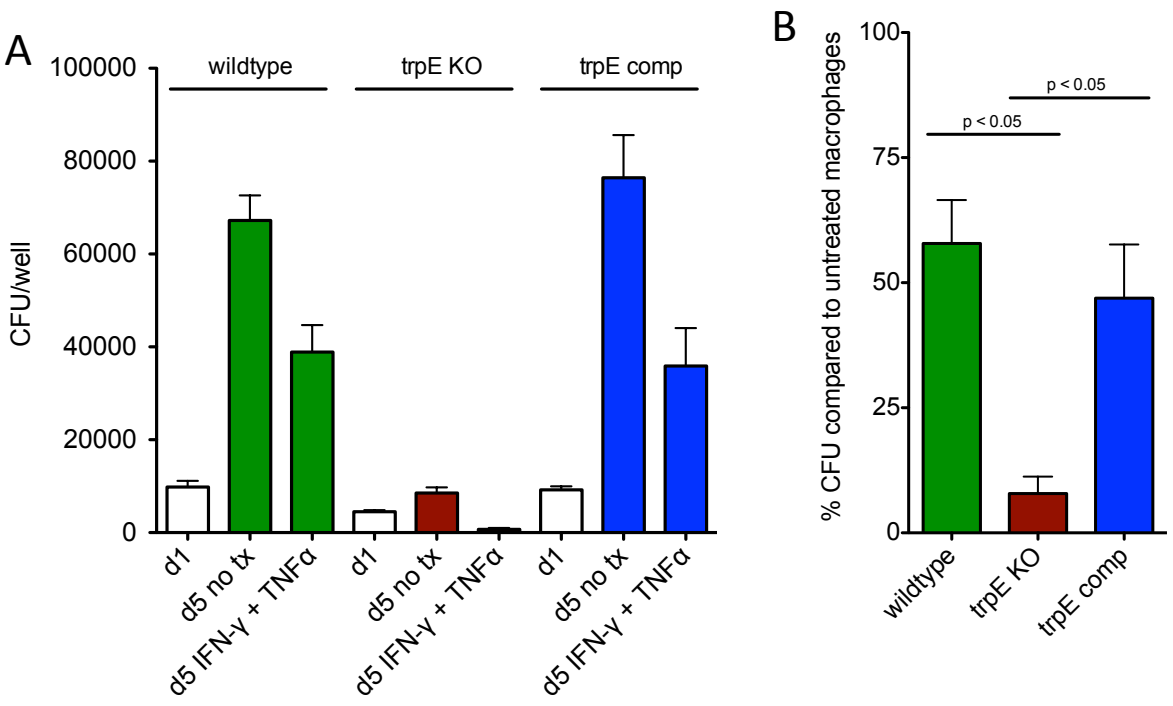


Figure 4.3. A.) Wildtype and *trpE* KO Mtb strains were infected into mouse peritoneal macrophages with and without IFN-γ and TNF-α. The approximate infectious dose was assessed by lysing the macrophages and plating for CFU at day 1 post-infection. Mycobacterial growth was assessed plating for CFU at day 5 post-infection. B.) The inhibitory effect of IFN-γ was measured by dividing the day 5 CFU from untreated macrophages by the CFU from IFN-γ treated macrophages.

IFN-γ induction of IDO necessitates mycobacterial tryptophan biosynthesis

To show that the auxotroph's hypersusceptibiliy to IFN-γ was tryptophan dependent, we added tryptophan to the media in IFN-γ-treated macrophages, and showed that the addition of tryptophan reversed the IFN-γ hypersuscpetbility (Fig 4.4A). As expected, tryptophan did not change the bacterial growth inhibitory effect of IFN-γ in wildtype Mtb, showing that tryptophan supplementation does not have a general growth effect on Mtb. Interestingly, tryptophan supplementation could not restore growth of the auxotroph in unstimulated macrophages to wildtype Mtb levels. It is possible that the levels of tryptophan needed to restore wildtype growth (1 mM in liquid broth) cannot be reached intracellularly, while the amount of tryptophan required to protect from IFN-γ-mediated killing could.

Many intracellular pathogens are natural tryptophan auxotrophs whose intracellular growth is also inhibited by IFN-γ stimulation, and we next attempted to determine if similar processes affected tryptophan availability in Mtb infection. One of the most highly transcriptionally induced genes in response to IFN-γ is a tryptophan-catabolizing enzyme, indoleamine-2,3-dioxygenase (IDO). IDO utilizes tryptophan as a synthetic precursor for kynurenines, immune signaling molecules that help control inflammation. In this synthetic process, it also greatly decreases the intracellular tryptophan pool. IDO is thus required for IFN-γ-mediated growth inhibition in *Chlamydia* and other tryptophan auxotrophic intracellular pathogens.

We tested the role of IDO in the *trpE* knockout's hypersusceptibility to CD4 T cell and IFN-γ by either a.) inhibiting IDO in both human and mouse macrophages with a specific chemical inhibitor, 1-methyl tryptophan (1-MT) (Fig 4.4B, Fig 4.5) or b.) using mouse macrophages derived from IDO knockout mice (Fig 4.4C)[13, 14]. In both cases, the hypersusceptibility was reversed. To demonstrate that 1-MT did not affect auxotroph growth, the trpE knockout was grown in 7H9 media with 1-MT. No growth was observed. This evidence demonstrates that CD4 T cells, likely through IFN-γ, depletes intracellular tryptophan, which forces Mtb to synthesize its own tryptophan.

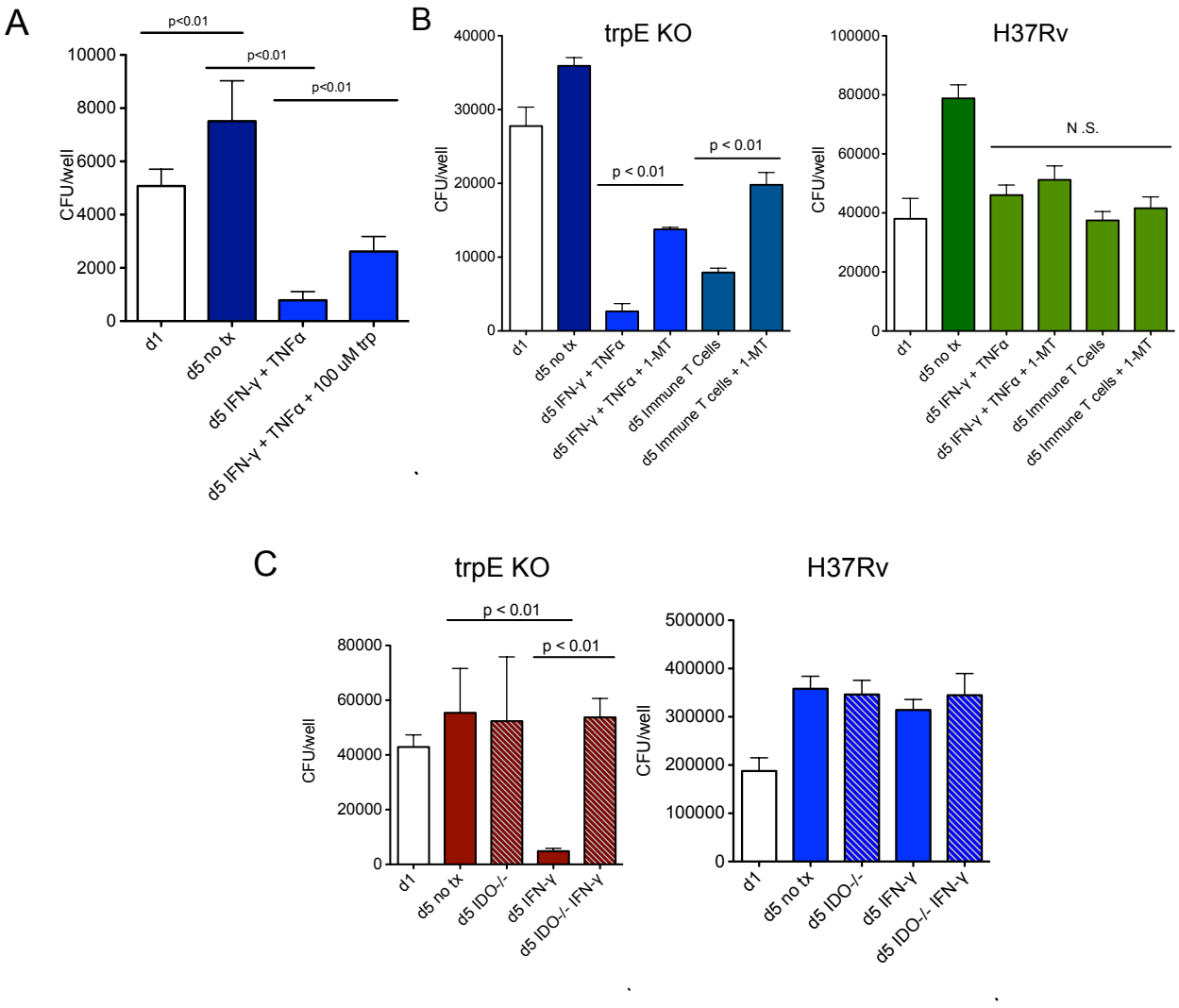


Figure 4.4. A.) *trpE* KO Mtb strains were infected into mouse peritoneal macrophages with and without IFN-γ and TNF-α. To determine if the cytokine effect was tryptophan dependent, 100 μM tryptophan was supplemented in the media. B.) Either IFN-γ and TNF-α or CD4 T cells from Mtb-infected mouse spleens were added on day 1 post-infection. The IDO inhibitor 1-MT was added to test the role of IDO in cytokine or CD4-mediated growth inhibition. C.) Macrophages were harvested from IDO knockout ($IDO^{-/-}$) mice, and infected with Mtb strains with and without IFN-γ and TNF-α.

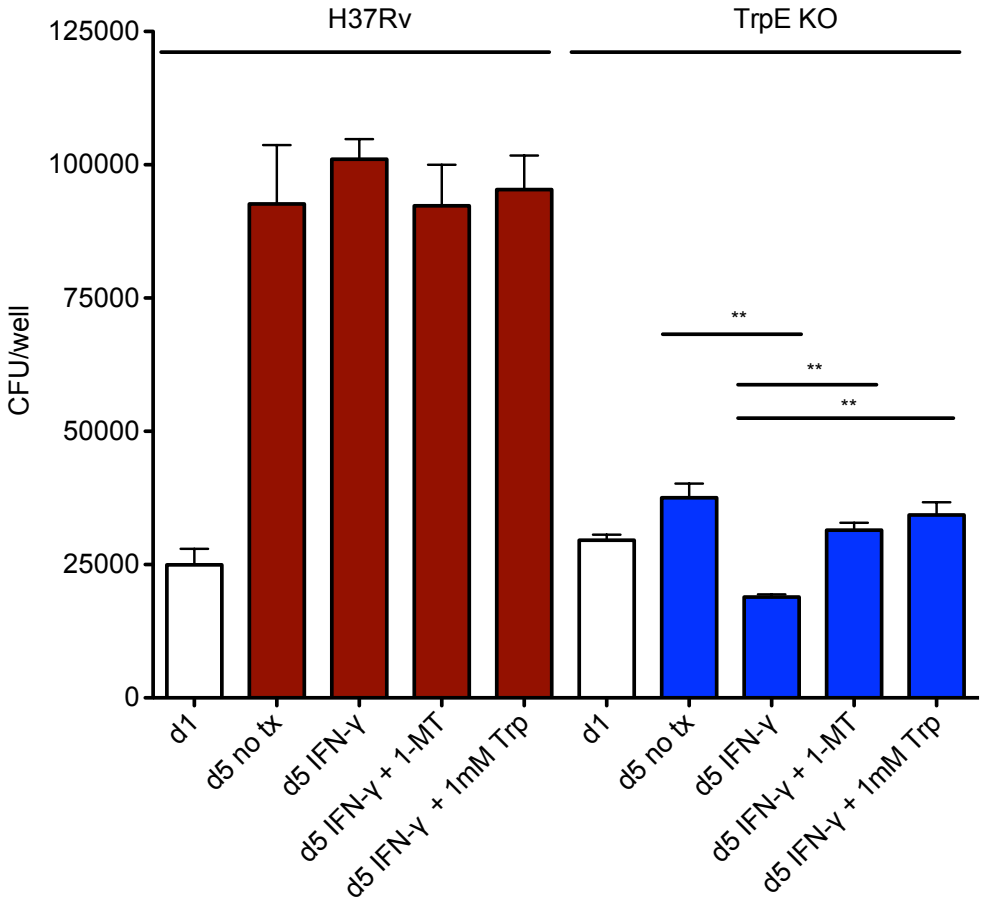


Figure 4.5. Human monocytes were isolated from buffy coats using magnetic CD14 antibodies. Cells were allowed to differentiate in the presence of GM-CSF for 5 days, and then infected with Mtb strains. Macrophages were further stimulated on day 1 post-infection with IFN-γ and treated with either 1-MT or supplemented with tryptophan.

Section 4.3: Anthranilate analogs kill Mtb in synergy with the immune system

Halogenated anthranilate analogs disrupt tryptophan biosynthesis to kill Mtb *in vitro*

The need for tryptophan biosynthesis in the face of CD4 immunity suggested that the pathway would be a good drug target *in vivo*. In other bacteria, anthranilate analogs have been used to poison anthranilate-utilizing synthetic pathways. *Pseudomonas aeruginosa* synthesizes quorum-sensing molecules called 4-Hydroxyl-2-alkylquinolones (HAQs) that are required for virulence[15]. Anthranilate is the synthetic precursor of HAQs, and three compounds in a series of halogenated anthranilates were able to block HAQ synthesis and decrease pathology of a *Pseudomonas* infection[15]. In the first committed step of tryptophan biosynthesis, TrpE converts chorismate to anthranilate, which is then converted to *N*-(5'-phosphoribosyl)-anthranilate by TrpD. Thus, we reasoned that anthranilate analogs might poison tryptophan biosynthesis as well, and we tested a panel of anthranilate analogs for Mtb growth inhibition in the presence and absence of tryptophan. Two fluorinated anthranilates, 2-amino-5-fluorobenzoic acid (5-FABA) and 2-amino-6-fluorobenzoic acid (6-FABA) had an MIC of 5 μM in liquid broth (Table 4.1 and Fig 4.6).

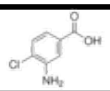
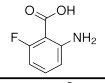
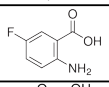
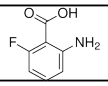
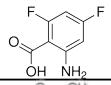
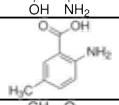
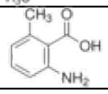
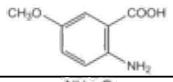
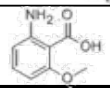
Compound	MIC in 7H9	MIC in 7H9 with 1 mM trp
4-CABA 	> 500 μM	> 500 μM
6-CABA 	> 500 μM	> 500 μM
4-FABA 	> 500 μM	> 500 μM
5-FABA 	5-10 μM	> 500 μM
6-FABA 	5-10 μM	> 500 μM
4,6-dFABA 	> 500 μM	> 500 μM
5-MeABA 	> 500 μM	> 500 μM
6-MeABA 	> 500 μM	> 500 μM
5-MeOABA 	> 500 μM	> 500 μM
6-MeOABA 	> 500 μM	> 500 μM

Table 4.1. MICs in 7H9 of Anthranilate analogs tested

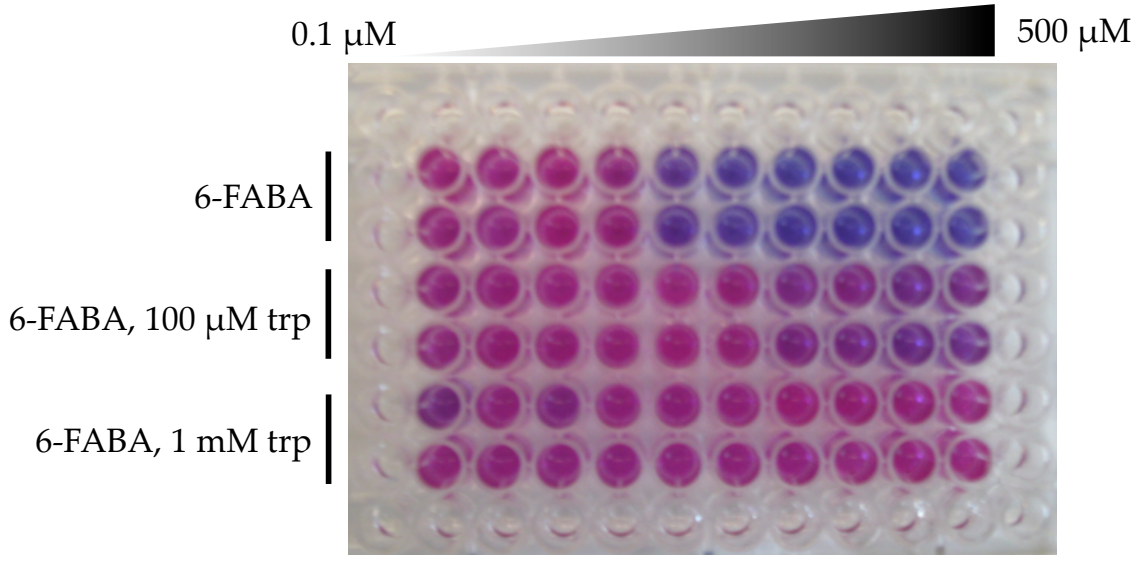


Figure 4.6. Alamar blue assay in 7H9 with and without tryptophan supplementation.

Neither compound had any measurable activity when the media was supplemented with 1 mM tryptophan, showing that the compounds affect tryptophan biosynthesis (Fig 4.7A). We also tested 5-FABA and 6-FABA activity in *Mycobacterium smegmatis (*M. smeg*)*. 5-FABA had an MIC of 10 μM in M. smeg, and 6-FABA had an MIC of 65 μM in M. smeg (Fig 4.8). Both molecules were rescued for growth in the presence of tryptophan, suggesting that the mechanism of action in M. smeg is similar to the mechanism in Mtb. Consistent with the auxotroph killing seen in tryptophan starvation, 6-FABA was also bactericidal in liquid broth (Fig 4.7B). Cultures were started in 7H9 broth with various amounts of 6-FABA, and bacteria were plated for CFU at days 0, 2, 3, and 6. At 20 μM 6-FABA, we saw a 100-fold decrease in CFU compared to the starting inoculum at day 6, and about a 10,000-fold decrease in CFU compared to the untreated control at the same time point.

To determine the exact mechanism of action of these compounds, we plated a total of 10^9 Mtb bacteria on 7H10 plates with 100 μM 6-FABA. About 300 colonies were retrieved, and 10 were picked and grown up in 100 μM 6-FABA in 7H9 broth. These colonies are currently being sequenced. Furthermore, strains overexpressing TrpE and TrpD were generated in Mtb and M. smeg. If either of those two enzymes are the targets of 6-FABA or its toxic metabolites, we would expect that overexpressors would be less susceptible to 6-FABA. These hypotheses are currently being tested.

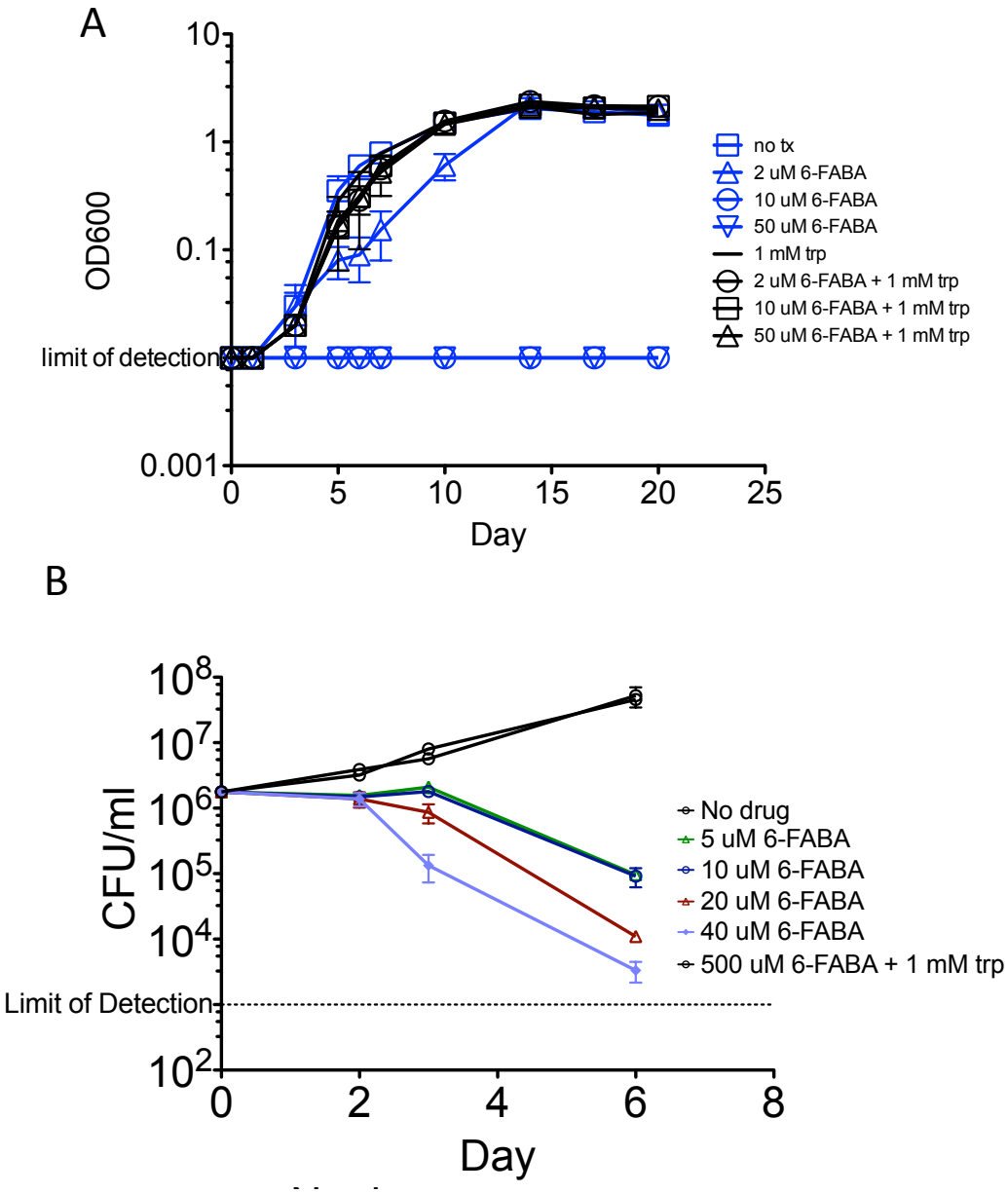


Figure 4.7. A.) Mtb was treated with 6-FABA and growth was assessed by measuring OD600. B.) Mtb was treated with 6-FABA and cultures were sampled and plated for CFU to test viability.

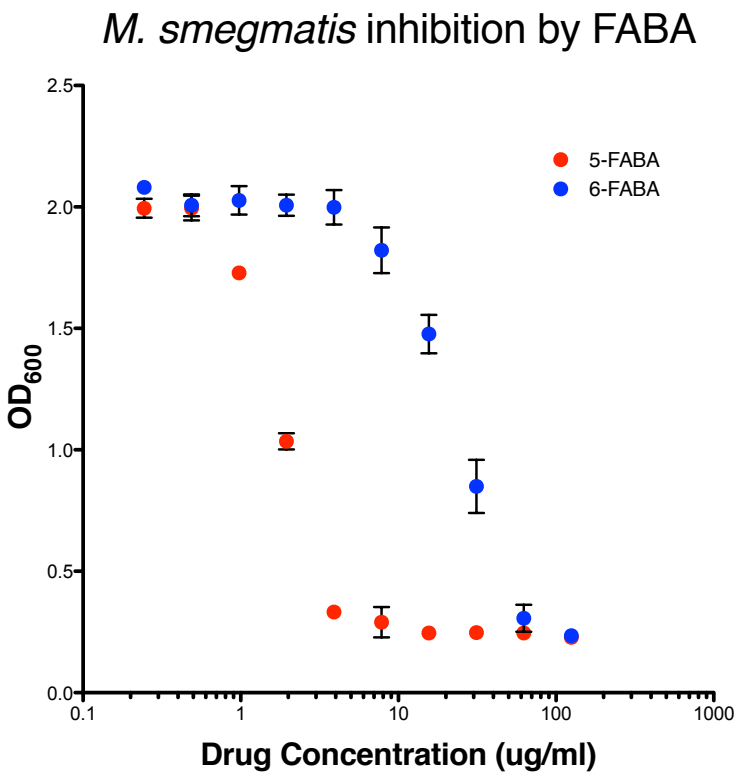


Figure 4.8. A.) Alamar blue testing for M. smeg MICs for 5-FABA and 6-FABA B.) M. smeg was grown in various amounts of 5-FABA and growth rates were measured (C.) to confirm the MIC *in vitro*.

6-FABA synergizes with IFN-γto kill Mtb in macrophages

We next tested 6-FABA in macrophages infected with Mtb. One day after infection, we added 6-FABA and measured bacterial growth by plating for CFU on day 5. At concentrations as low as 10 μM, 6-FABA significantly limited growth by over 10-fold (Fig 4.9A). The number of CFU at d5 was lower than at d1, suggesting that 6-FABA had bactericidal activity in macrophages. Chemical inhibitors that affect macrophage survival can also decrease bacterial growth, so we performed viability tests to exclude that as a possibility. LDH release did not increase with 6-FABA treatment of macrophages, demonstrating that 6-FABA is not cytotoxic (Fig 4.9B).

Since the tryptophan auxotroph was hypersusceptible to the effects of IFN-γ stimulation, we hypothesized that 6-FABA's block of tryptophan biosynthesis would work in synergy with IFN-γ to kill Mtb in macrophages. To test this, we dosed both IFN-γ (10U/ml) and 6-FABA (0.2 μM) to a level where each individually had about a 2-fold inhibitory effect on bacterial growth (Fig 4.10A and C). Without synergy, we predicted that the combined effect of 6-FABA and IFN-γ would be about 4-fold. Instead, the effect was 41.7-fold in mouse and 8.7-fold in human macrophages, demonstrating clear synergy (Fig 4.10B and D).

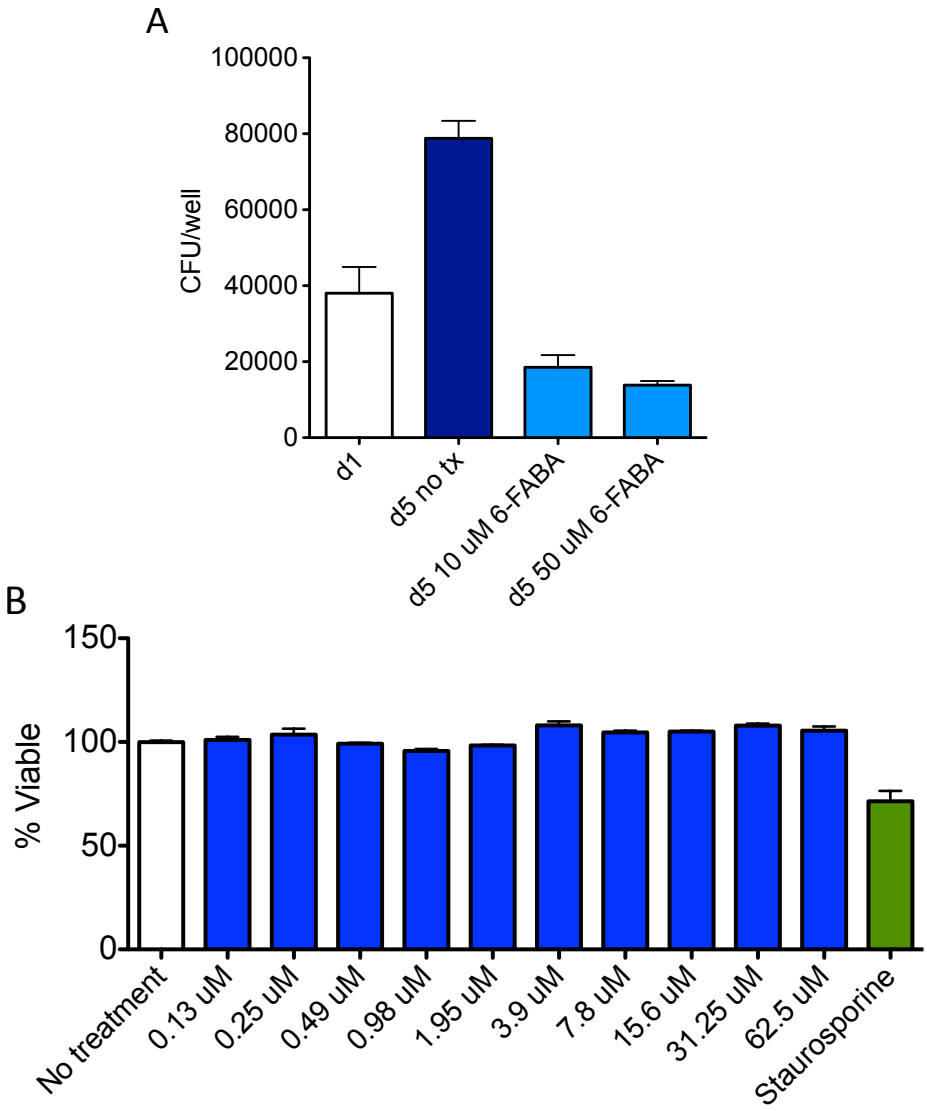


Figure 4.9. A.) Mouse peritoneal macrophages were infected with Mtb. On day 1 post-infection, macrophages were treated with 6-FABA. Cells were lysed and plated for CFU on day 5. B.) Mouse peritoneal macrophages were treated with increasing doses of 6-FABA or staurosporine, and viability was calculated using an assay for LDH release.

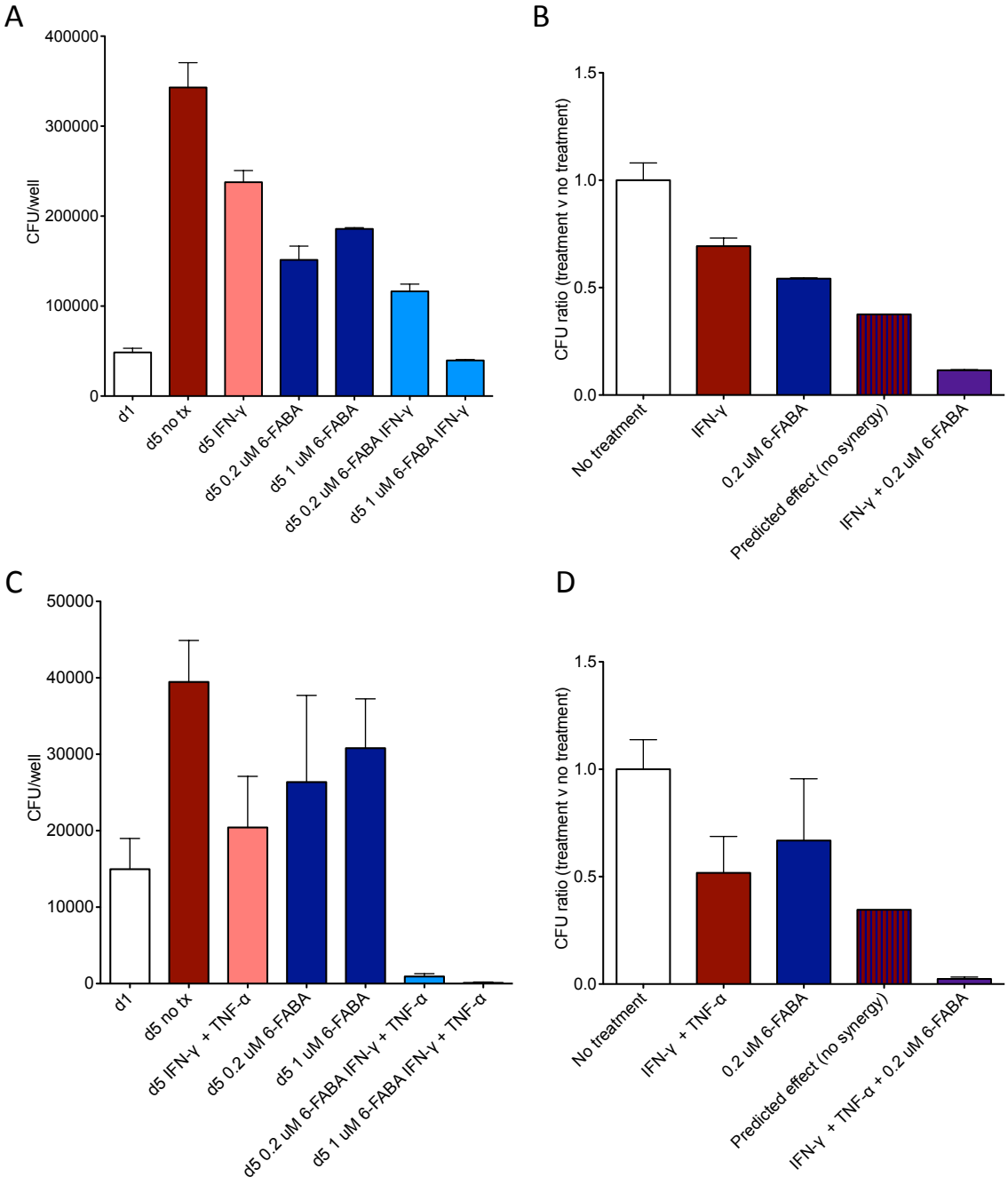


Figure 4.10. A.) Human monocyte-derive and C.) mouse peritoneal macrophages were infected with Mtb and treated with IFN-γ (with vitamin D in human macrophages and with TNF-α in mouse macrophages) and/or 6-FABA on day 1. Doses were titrated such that the effect of a single treatment would decrease CFU compared to the untreated macrophage control. CFU ratios of the experimental conditions to the untreated controls were calculated for both human (B.) and mouse (D.) experiments. Using these ratios, the predicted non-synergistic value was determined and compared to the actual combined effect of IFN-γ and 6-FABA.

6-FABA is bioavailable when orally dosed in mice.

Synergy with IFN-γ in macrophages suggested that 6-FABA could kill Mtb *in vivo* in conjunction with CD4 T cells. Since acids are often poorly tolerated and absorbed in oral formulations, we synthesized an ester derivative in hopes of improving pharmacokinetic and tolerability profiles in mice. Since the ester had no anti-mycobacterial activity *in vitro*, we relied on esterase activity *in vivo* to generate the active-form acid. Using both 6-FABA and the ester derivative, we first determined that both compounds were bioavailable in with both oral and intraperitoneal delivery (Fig 4.11A and B), and that oral doses of up to 250 mg/kg/day did not result in clinical illness or weight loss. 5-FABA was not as bioavailable as 6-FABA in mice (Fig 4.11A and B). While the ester was not detectable in plasma of treated mice, the acid form was detectable at almost identical amounts to the 6-FABA-dosed mice, suggesting that the ester was rapidly converted to 6-FABA *in vivo*.

We infected mice with 10^2 aerosolized Mtb bacilli, and allowed infection to establish in the lungs for 8 days. We treated mice six times a week with either INH (25 mg/kg/day), 6-FABA (200 mg/kg/day) or the ester derivative (200 mg/kg/day), and will measure bacterial growth 2 weeks and 4 weeks after treatment initiation. This experiment is currently ongoing. The pharmacokinetic properties measured predict that the plasma levels of 6-FABA in both 6-FABA and ester derivative dosed mice will stay above the in vitro MIC for most of the

treatment duration. This is an encouraging fact, and we look forward to measuring the efficacy of these compounds *in vivo*.

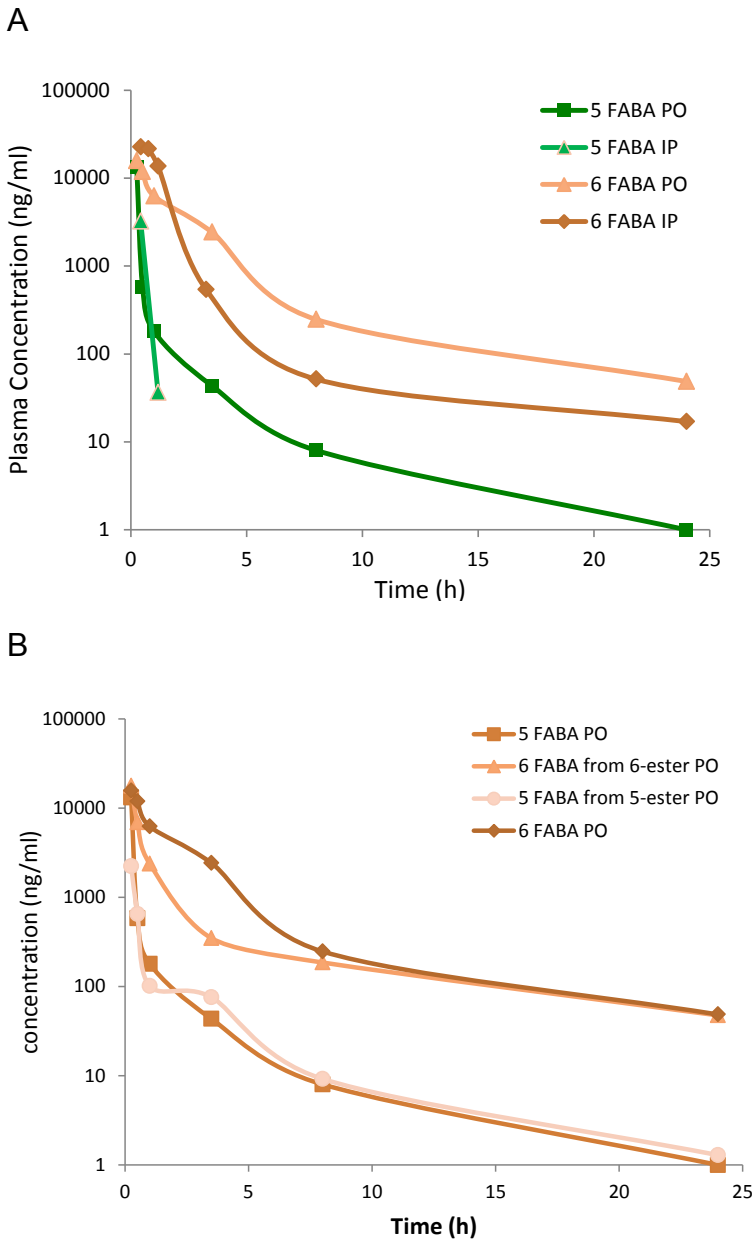


Figure 4.11. A.) 5-FABA and 6-FABA (25 mg/kg) were given at a single dose by intraperitoneal injection or oral gavage. At 15 minutes, 1 hour, 4 hours, 8 hours and 24 hours, blood was drawn and levels of 5-FABA and 6-FABA were measured in the plasma. B.) For 5-fluoro and 6-fluoro derivatives of anthranilate, both ester and acid forms were delivered by oral gavage and acid levels were measured in the plasma. Ester levels were not detectable.

Discussion

In this study, the high-throughput profiling in Chapter 3 helped us narrow our search for candidate virulence factor drug targets. Targeting virulence factors has been touted recently for many reasons, including i.) expansion of the potential target list beyond *in vitro* essential genes, ii.) protection for the host against virulence factor toxicity, and iii.) reduction of resistance development for drugs whose targets do not affect viability[16-18]. By searching for mycobacterial processes that protect Mtb from CD4 T cell-mediated killing, we hoped to identify drug targets whose inhibition made Mtb more susceptible to antimycobacterial immunity. Since tryptophan auxotrophs, which could not survive in wt mice, were rescued in the absence of CD4 T cells, we determined that blocking tryptophan biosynthesis would allow CD4 T cells to kill Mtb.

Small molecule inhibitors of tryptophan and other aromatic amino acid biosynthesis exist for other organisms, suggesting that certain enzymes in these pathways might be "druggable," or suitable for chemical inhibition[19]. Furthermore, tryptophan is an essential amino acid for humans[20, 21]. Since hosts get all of their needed tryptophan from their diets, any cross-species activity of anti-tryptophan synthesis drugs should be non-toxic. This was experimentally validated by the lack of toxicity of 6-FABA in both cultured macrophages and treated mice. Halogenated anthranilates were first described as inhibitors of a *Pseudomonas aureginosa* quorum-sensing pathway[15]. The same group also noted that 6-FABA inhibited *in vitro P. aureginosa* growth in the absence of tryptophan. After testing

a series of anthranilate analogs, we found that both 6-FABA and 5-FABA had bactericidal activity against Mtb at low concentrations, though only in the absence of tryptophan.

Rescue of growth with tryptophan strongly suggests that 6-FABA affects the tryptophan biosynthesis pathway, but it does not prove that 6-FABA inhibits tryptophan production. Because TrpD utilizes anthranilate as a substrate, we speculated that 6-FABA might be a competitive inhibitor of TrpD. However, it is also possible that 6-FABA could be a substrate of TrpD, resulting in the production of fluorinated intermediates that inhibit tryptophan biosynthesis or even fluorinated tryptophan that incorporates into and poisons proteins. In both these cases, the addition of exogenous tryptophan could conceivably rescue the toxic effect.

Until now, the most promising work in TB drug development has produced small molecule inhibitors of a limited number of cell division and growth mechanisms[22-26]. In order to expand our medical arsenal against TB, both new drugs and new targets are needed. Mtb has specifically evolved to survive in an infectious niche, a complex *in vivo* environment created by the interplay between host immunity and pathogen virulence. Here, we demonstrate that these virulence determinants, the mycobacterial products that allow Mtb survive host defenses, can be effective drug targets. In immunocompetent individuals, Mtb infection results in a complex interplay between host and pathogen that can result in disease. By targeting

tryptophan biosynthesis in this study, and hopefully other virulence factors in the future, we make Mtb more susceptible to immunity and tip the balance in favor of host.

Section 4.4: Materials and Methods

Construction of trpE knockout and complement strains

The hygromycin resistance gene was amplified with flanks containing 500 bp regions upstream and downstream (and slightly overlapping with) of trpE (See Table 4.2 for primers). This construct was electroporated into a strain of Mtb containing the plasmid pNIT(kan)::RecET-SacB, which contains the machinery for mycobacterial recombineering. Recombineering-competence was achieved by induction of the RecET complex by the addition of 1 mM isovaleronitrile (IVN, Sigma Aldrich) to a culture at OD600 0.8. IVN addition induced expression of the recombineering machinery on pNit(kan)::RecET-SacB. After 8h, 10 mL of 2M glycine was added, and the culture was grown overnight to yield recombineering-competent Mtb. Transformations were plated on 7H10 agar with 1 mM tryptophan. Positive clones were plated on 7H10 agar containing 10% sucrose to counterselect against the recombinase plasmid, and scored for growth on kanamycin-containing agar to confirm the loss of pNit::RecET. Deletion of the endogenous locus was confirmed by PCR and by phenotypic tests for auxotrophy. Finally, to complement, the *trpE-rv1610* two-gene operon was amplified and cloned, along with an artificial promoter, using multisite gateway into pDE43-MCK, which integrates into the L5 site (see Table 4.2 for primers).

Macrophage infections

Peritoneal macrophages were stimulated with Thioglycollate Medium (3%) by intraperitoneal infection. Three – five days after injecting, the peritoneum was exposed and macrophages were harvested by adding 10 mL of RPMI media with

penicillin and streptomycin into the peritoneal cavity and retrieving the fluid. Cells were washed and incubated with CD11b microbeads (Miltenyi) for 15 minutes on ice. Magnetic separation of CD11b+ cells was completed, and cells were plated for 2 days before use. Human monocyte-derived macrophages were made from donated buffy coats. Buffy coats were diluted and layered on Ficoll for gradient centrifugation. The middle cellular layer was harvested and washed multiple times, and then incubated with CD14 microbeads (Miltenyi) on ice for 15 minutes for magnetic isolation. Cells were then grown in GM-CSF (10 ng/ml) for 5 days before use. Cells were infected with Mtb at an MOI of 10:1 for 2 hours at 37° C. They were then washed 4 times with RPMI media and grown overnight. On day 1 post-infection, most conditions were added. To lyse cells, Triton-X 100 (0.1%) in PBS was added to each well. Serial dilutions of the lysis were made and plated.

Alamar Blue Assays

In a 96-well plate, bacterial cultures were started at an OD600 of 0.003. After five days, resazurin reagent was added and cells were kept in a shaking 37° C incubator for another 1-2 days. Plates were read and the MIC was determined as the first concentration at which the color changes.

Table 4.2. Primers

Primer Name	Primer Sequence	Notes
JZ.140 trpE F upstream	GCG AAT TTG GCC TGC TTT C	Primer used to amplify the 5' end of the 5' flank for trpE knockout
JZ.160 trpE R UP-ST XTRA	ctaCTTGTCGTCGTCGTCCTTCTG TCG GCC AAG ACC TTG	Primer used to amplify the 3' end of the 5' flank for trpE knockout
JZ.161 trpE F DN-ST XTRA	GTTTTTTTGGGCCTAGGGAAGGACA ACG GAT CCT ACG AAT ACA AC	Primer used to amplify the 5' end of the 3' flank for trpE knockout
JZ.143 trpE R downstream	CGT CAC CAC TGG GAC ATG	Primer used to amplify the 3' end of the 3' flank for trpE knockout
JZ.156 HygF lox XTRA	AAGGACGACGACGACAAGtagCGCTCTAG AACTAGTGGATCC	Primer used to amplify the hyrgromycin cassette with stitch PCR sequences.
JZ.157 HygR lox XTRA	TCCTTCCCTAGGCCCAAAAAAACTCT AGA CTC GAG GTA CCG G	Primer used to amplify the hyrgromycin cassette with stitch PCR sequences.

Section 4.5: Bibliography

1. Byrne, G.I., Lehmann, L.K. & Landry, G.J. Induction of tryptophan catabolism is the mechanism for gamma-interferon-mediated inhibition of intracellular Chlamydia psittaci replication in T24 cells. *Infection and immunity* 53, 347-351 (1986).

2. Murray, H.W. et al. Role of tryptophan degradation in respiratory burst-independent antimicrobial activity of gamma interferon-stimulated human macrophages. *Infection and immunity* 57, 845-849 (1989).

3. Beatty, W.L., Belanger, T.A., Desai, A.A., Morrison, R.P. & Byrne, G.I. Tryptophan depletion as a mechanism of gamma interferon-mediated chlamydial persistence. *Infection and immunity* 62, 3705-3711 (1994).

4. Daubener, W. et al. Restriction of Toxoplasma gondii growth in human brain microvascular endothelial cells by activation of indoleamine 2,3-dioxygenase. *Infect Immun* 69, 6527-31 (2001).

5. Silva, N.M. et al. Expression of indoleamine 2,3-dioxygenase, tryptophan degradation, and kynurenine formation during in vivo infection with Toxoplasma gondii: induction by endogenous gamma interferon and requirement of interferon regulatory factor 1. *Infection and immunity* 70, 859-868 (2002).

6. Medzhitov, R. et al. Highlights of 10 years of immunology in Nature Reviews Immunology. *Nat Rev Immunol* 11, 693-702 (2011).

7. Ibana, J.A. et al. Inhibition of indoleamine 2,3-dioxygenase activity by levo-1-methyl tryptophan blocks gamma interferon-induced Chlamydia trachomatis persistence in human epithelial cells. *Infect Immun* 79, 4425-37 (2011).

8. Leonhardt, R.M., Lee, S.J., Kavathas, P.B. & Cresswell, P. Severe tryptophan starvation blocks onset of conventional persistence and reduces reactivation of Chlamydia trachomatis. *Infect Immun* 75, 5105-17 (2007).

9. Parish, T. Starvation survival response of Mycobacterium tuberculosis. *Journal of Bacteriology* 185, 6702-6706 (2003).

10. Smith, D.A., Parish, T., Stoker, N.G. & Bancroft, G.J. Characterization of auxotrophic mutants of Mycobacterium tuberculosis and their potential as vaccine candidates. *Infection and immunity* 69, 1142-1150 (2001).

11. Russell, D.G. Who puts the tubercle in tuberculosis? *Nat Rev Microbiol* 5, 39-47 (2007).

12. Ernst, J.D. The immunological life cycle of tuberculosis. *Nat Rev Immunol* 12, 581-91 (2012).

13. Baban, B. et al. Indoleamine 2,3-dioxygenase expression is restricted to fetal trophoblast giant cells during murine gestation and is maternal genome specific. *J Reprod Immunol* 61, 67-77 (2004).

14. Belladonna, M.L. et al. Kynurenine Pathway Enzymes in Dendritic Cells Initiate Tolerogenesis in the Absence of Functional IDO. *Journal of immunology (Baltimore, Md : 1950)* 177, 130-137 (2006).

15. Lesic, B. et al. Inhibitors of pathogen intercellular signals as selective anti-infective compounds. *PLoS Pathog* 3, 1229-39 (2007).

16. Barczak, A.K. & Hung, D.T. Productive steps toward an antimicrobial targeting virulence. *Curr Opin Microbiol* 12, 490-6 (2009).

17. Clatworthy, A.E., Pierson, E. & Hung, D.T. Targeting virulence: a new paradigm for antimicrobial therapy. *Nat Chem Biol* 3, 541-8 (2007).

18. Rasko, D.A. & Sperandio, V. Anti-virulence strategies to combat bacteria-mediated disease. *Nat Rev Drug Discov* 9, 117-28 (2010).

19. Coggins, J.R. et al. Experiences with the shikimate-pathway enzymes as targets for rational drug design. *Biochemical Society Transactions* 31 (2003).

20. WHO. Protein and Amino Acid Requirements in Human Nutrition. (2007).

21. John, A. & Bell, J.M. Amino Acid Requirements of the Growing Mouse. *The Journal of Nutrition* 106, 1361-1367 (1976).

22. Zumla, A., Hafner, R., Lienhardt, C., Hoelscher, M. & Nunn, A. Advancing the development of tuberculosis therapy. *Nat Rev Drug Discov* 11, 171-2 (2012).

23. Stover, C.K. et al. A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis. *Nature* 205, 962-966 (2000).

24. Andries, K. et al. A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. *Science* 307, 223-7 (2005).

25. Villemagne, B. et al. Tuberculosis: the drug development pipeline at a glance. *Eur J Med Chem* 51, 1-16 (2012).

26. Makarov, V. et al. Benzothiazinones kill Mycobacterium tuberculosis by blocking arabinan synthesis. *Science* 324, 801-4 (2009).

Chapter 5.

Summary and Perspectives.

Section 5.1: Summary of Findings

Introduction

Defining essential genes has traditionally been thought of as an important first step in drug design. The argument states that essential genes can serve as a high-pass filter for selecting protein targets of future therapeutics. Further considerations, both biolgical (vulnerability of a target as defined by the amount of inhibition that results in bacterial death, propensity of resistance development) and chemical (size and polarity of potential binding pocket, structural similarity to previously targeted proteins) help to determine the promise of the protein as a therapeutic target[1-5]. Essential genes are traditionally determined as the genes required for growth in laboratory culture conditions[6-8]. However, Mtb replicates primarily in the human lung, which is a very different environment[9]. Effective drugs need to inhibit Mtb replication *in vivo*. Many genes that are required for replciation *in vivo* are not required for growth *in vitro*, and it is possible that certain genes required for growth in laboratory cultures and dispensible for bacterial growth during infection. Thus the search for *in vitro* essential genes is a good start, but it is both limited in scope and potentially misleading as a method of defining good *in vivo* drug targets. Furthermore, the immune system imposes a diverse set of pressures on the bacterium, and bacterial survival is dependent on multiple factors to survive these pressures[10]. By understanding these particular determinants of survival, we can leverage immune-imposed stresses to kill Mtb more effectively. Thus understanding the environment under which Mtb

replicates and the adaptive tools Mtb possesses allows us to define new and previously underexplored drug targets.

To that end, the narrative in this dissertation is told in three parts. In Chapter 2, I present an improved methodology to define essential genomic regions in Mtb. The method—which involves the direct sequencing of transposon junctions in a transposon library—gives us a new picture of genetic requirement with much higher resolution and accuracy. In Chapter 3, I present our search for genetic requirement in many models relevant towards human disease. These include *in vitro* conditions that mimic immune-mediated stresses, including amino acid starvation, nitrosative stress, acid stress and iron depletion. Most importantly, I present the genetic requirements for survival during a mouse infection, and the specific determinants for surviving the CD4 T cell response. The latter determinants were particularly interesting to us. We reasoned that many of these genes could help us identify attempted (but failed) methods of CD4-mediated killing. By inhibiting the survival determinants, however, we predicted that we could turn these attempted killing strategies into effective mechanisms of immune control. To that end, Chapter 4 focuses on one of these attempted killing-survival strategy pairs. We show that CD4 T cells attempt to kill Mtb by starving the bacteria of tryptophan, and delineate the pathway by which this occurs. We also present a new inhibitor of bacterial tryptophan biosynthesis and show that the inhibitor is able to kill Mtb in synergy with adaptive immunity.

Defining Essential Genes in Mtb

To define genes required for optimal growth *in vitro*, we introduced transposons to a pool of bacteria, creating a library of mutants in which each bacterium had a single transposon insertion. Coupled with the assumption that essential genes will not sustain insertions, this strategy has been used in Mtb and other bacteria for over a decade to define essential genes[8, 11-13]. When we started this work, deep sequencing was becoming increasingly accessible and inexpensive[14]. Along with Chris Sassetti's lab, we helepd develop a platform in Mtb that specifically sequenced transposon-junctions[6, 7, 15]. While sequencing transposon junctions in other bacterial organisms was being developed concurrently, we realized that by mapping the exact location of transposon insertions, we had the resolution to determine genetic requirement beyond the level of the gene[16-19]. We took two approaches to leverage this resolution. First, we used unbiased sliding windows to define required regions irrespective of annotated features[7]. Second, we used an extreme value distribution to define essential stretches of insetion deserts, irrespective of the overall insertional landscape in the rest of the gene[6]. These methods allowed us to independently call essential genes during *in vitro* growth.

Defining Genes Required for Growth in Infection Models

Normal laboratory media does a poor job of modeling the *in vivo* environment of Mtb, so we sought to expand our genetic requirement analysis to models that mimic the more stressful milieu of the infected and inflammed human lung. We did this *in vitro* by passing our library through a number of conditions that

estimate certain *in vivo* stresses. We also modeled disease environment by infecting our library into mice. While the screen for genes required for growth in vivo yielded the most hits, as well as the hits with the largest effects, it was impossible to predict what elements of mouse host defenses each gene was responsible for combating. We reasoned we could start to tease apart the components of host immunity and their "cognate" bacterial *in vivo* survival determinants by doing the same screen in immune deficient mice. If genes were specifically required for combating a certain aspect of mouse immunity, these genes would be necessary for growth in wildtype mice but dispensible in mice lacking these immune components. There were many to choose from, including mice lacking the most critical components for mycobacterial immunity such as IFN-γ, TNF-α and iNOS[20-22].

We chose to focus on the CD4 T cell response because of two main factors. First, it is clearly relevant in both mouse and human immunity against Mtb, as CD4-depleted mice and HIV-infected humans do much worse with TB compared to their immunocompetent counterparts[23-28]. Second, while we knew that CD4 T cells were crucial, the nature of the stress they imposed was, and still is, largely uncharacterized. We hoped that by defining the genes required for combating CD4 T cells, we would not only identify interesting bacterial survival determinants with future drug-target potential, but also use bacterial requirement profiles as a bioprobe to describe the nature of CD4 stress. The results of the screen had multiple facets, and the rest of the dissertation follows one particular result that

suggested a mechanism of CD4-mediated tryptophan starvation and bacterial adaptation through tryptophan biosynthesis. I followed one other pathway, gluconeogenesis, but the mechanism was difficult to determine. I will discuss some other hits below in the discussion of perspectives.

Determining the mechanism of tryptophan starvation

Tryptophan starvation through IDO is a well-described phenomenon in cell culture models of intracellular pathogens. IFN-γ induces IDO, which catabolizes tryptophan in macrophages and kills natural tryptophan auxotrophs[29-31]. We showed that the same pathway is in play during Mtb infection, and that in this case it involves CD4 T cells. The tryptophan auxotroph Mtb strain was hypersusceptible to the effects of CD4 T cells, and this hypersusceptibility was IDO-dependent and reversible by the addition of tryptophan. Unlike *Chlamydia*, though, Mtb is able to survive this host defense by synthesizing its own tryptophan. By elucidating the effect of this pathway on Mtb and Mtb's response, we uncovered a vulnerable point in the Mtb lifecycle. CD4 T cells attempt to kill the bacteria by tryptophan starvation and as long as bacterial tryptophan biosynthesis is interrupted, this attempt works as a method of controlling Mtb growth.

Discovering a chemical inhibitor of tryptophan biosynthesis

By combining exogenous tryptophan starvation by CD4 T cells with endogenous tryptophan biosynthesis block by genetic manipulation, we created synergy that

induced profound bacterial killing in macrophages. This encouraged us to find a chemical inhibitor of tryptophan biosynthesis as a potential drug that worked in synergy with CD4 T cells. The fluorinated anthranilates constituted a small series of compounds that we could easily test for tryptophan biosynthesis blockade. Two of these molecules, 6-FABA and 5-FABA, had low MICs in Mtb, with bactericidal effects *in vitro* that were reversible with tryptophan. The molecules killed Mtb in synergy with IFN-γ in macrophages, and have good pharmacokinetic properties in mice. We are eagerly awaiting the results from an ongoing mouse efficacy study.

This work demonstrates that there might yet be some promise in rationale drug design. We saw the life cycle of Mtb in the mouse model as a series of stress-inducing components, each presenting a vulnerable checkpoint for the bacteria. Mtb survives because it possesses the tools to pass these checkpoints. For nitrosative stress, it has the proteasome to tidy up damaged proteins and nucleotide excision repair to fix damaged DNA[32, 33]. For changes in carbon sources it adapts by switching metabolic activity to fatty acid catabolism and gluconeogenesis[15, 34-37]. While Mtb successfully overcomes these checkpoints, they still represent a point of potential vulnerability, since inhibition of these checkpoint survival determinants restores the efficacy of the checkpoints. Host-pathogen interactions are characterized by this give and take between the immune system and the pathogen. These battles have been fought for many evolutionary generations, and our immune system possesses many ancient

antimicrobial pathways that work against certain bacteria but not Mtb. By deciphering how Mtb survives these pathways, we can revive these ineffective pathways and turn them back into bacteria-killing host defenses.

Section 5.2: Perspectives

The data mouse transposon screen remains a largely unmined bed of useful information. As a matter of volume, there are hundreds of uncharacterized genes that have been newly predicated as essential for growth *in vivo*. Traditional bacterial genetics can help define the activity of many of these hits and the reason for their *in vivo* requirement. Interesting pathways that were uncovered as virulence factors include a series of four thiosulfate transferases, a membrane transporter and ATPase (*ftsEX*) known in other bacteria to regulate cell division in specific circumstances (sporulation, low-osmotic culture conditions), the synthetic pathway (*fbiABC*) for the F420 coenzyme, and the alternative, or bc, cytochrome (*cydABCD*).

The result of this screen lends itself well to candidate approaches for understanding bacterial physiology during infection, especially for pathways that are hit multiple times or hit in their entirety in the screen. However, a majority of our hits do not fall into this category. Not only are they not predicted to be in known pathways, but they do not have predicted functions at all. To start probing the function of these genes, we can start to employ more hypothesis-generating bioinformatic tools. One is the gene set overlap analysis done in Chapter 3 of this thesis. By overlapping the gene sets required for different conditions, we not only predict the nature of certain stresses, but we also create signatures of requirement for genes of unknown function. Requirement signatures can also be

overlapped with transcriptional signatures. In particular, transcriptional signatures can be used to create networks of coregulated genes, and searching for networks that are enriched in requirement gene sets can also help probe both the biological function of genes of unknown function and the environmental milieu of uncharacterized niches.

The method applied to searching for CD4-survival genes can also be applied to many other immune deficient mice. A previous graduate student characterized mutants that were rescued as well as mutants that were hypersusceptible in iNOS deficient mice[38]. With higher resolution and higher confidence tools, it is worth trying this experiment again. Another particularly interesting model could be diabetic mice. It is unclear why diabetes increases TB risk, and assessing the genetic requirements for survival in a diabetic mouse could help shed light on the key determinants of growth-permissiveness in diabetes. Mice are not the best model of human disease, since they do not form human-like granulomas, do not seem to be able to ever decrease bacterial populations, and invariably die of TB infection. Macaques are also imperfect, but they do a much better job of modeling both the granuloma structure and the differential clinical outcomes of human infection and disease. A tranpsoson library screen seems increasingly possible in macaques; when done, it will certainly be informative. Comparisons between genetic requirements of bacterial survival in a macaque versus a mouse will be fascinating, and perhaps describe in more detail the differences between the two models.

The key motivation for our work is the breadth and depth of human suffering caused by TB around the world. The numbers of infected individuals are staggering: 2 billion infected, 2 million deaths per year, 10 million new cases each year, 500,000 MDR cases each year[39]. But the toll of TB is more clearly understood at the community level. Of the other major killers worldwide, many affect primarily older individuals (cancer, cardiovascular diseases) or young children (diarrheal diseases). TB, especially in the world of growing HIV prevalence, affects primarily working-age adults. From a socioeconomic perspective, this is catastrophic. Poor communities, those most affected by the HIV and TB epidemics, are losing the most productive parts of their societies. From a human perspective, the loss is much more difficult to calculate.

TB also has a habit of targeting the most vulnerable. In rich-world nations like ours, it was New York prison inmates in the 1990s, and homeless populations and poor immigrant families today[40]. In Russia, the overburdened and corrupt criminal justice system that retains people in their overrun prison system longer than they should, or than they mean to, is the epicenter of TB transmission[41]. In China, TB is a reemerging epidemic amongst migrant worker populations flooding the cities for work as rural development takes a back seat to the government's urban growth aspirations[42]. TB is both poverty-creator and poverty-consequence. Either way, it is inextricably linked to those in our world who are underserved, underresourced, and whose lives too often fly under the radar. Our

efforts at understanding the biology of TB necessarily falls within this global context. In the last data chapter of this thesis, we make a small attempt to leverage our biological understanding to develop a potential new therapy against TB. It is a drop in the bucket, and thankfully so. There are many promising potential therapeutics in clinical trials already, and the TB drug development pipeline must not go dry, since drug resistance will be a constant battle for all new drugs. Public health interventions continue to roll out, and the extension of DOTS into new communities will make huge inroads in limiting the epidemic. Efforts need to be made on all levels to translate what knowledge we have to create effective treatments, preventions, and programs for TB. With the right allocation of resources to the most vulnerable populations, we have the tools now to stamp out the TB epidemic[43]; and with the development of better tools—more effective vaccines, quicker-acting drugs—the goal of TB eradication will inch ever closer.

Section 5.3: Bibliography

1. Wei, J. et al. Depletion of antibiotic targets has widely varying effects on growth. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 4176-4181 (2011).

2. Makarov, V. et al. Benzothiazinones kill Mycobacterium tuberculosis by blocking arabinan synthesis. *Science* **324**, 801-4 (2009).

3. Payne, D.J., Gwynn, M.N., Holmes, D.J. & Pompliano, D.L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **6**, 29-40 (2007).

4. Villemagne, B. et al. Tuberculosis: the drug development pipeline at a glance. *Eur J Med Chem* **51**, 1-16 (2012).

5. Sacchettini, J.C., Rubin, E.J. & Freundlich, J.S. Drugs versus bugs: in pursuit of the persistent predator Mycobacterium tuberculosis. *Nat Rev Microbiol* **6**, 41-52 (2008).

6. DeJesus, M. et al. Bayesian Analysis of Gene Essentiality based on Sequencing of Transposon Insertion Libraries. *Bioinformatics* **10**, 1093 (2013).

7. Zhang, Y.J. et al. Global assessment of genomic regions required for growth in Mycobacterium tuberculosis. *PLoS Pathog* **8**, e1002946 (2012).

8. Sassetti, C.M., Boyd, D.H. & Rubin, E.J. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular microbiology* **48**, 77-84 (2003).

9. Ernst, J.D. The immunological life cycle of tuberculosis. *Nat Rev Immunol* **12**, 581-91 (2012).

10. Ehrt, S. & Schnappinger, D. Mycobacterial survival strategies in the phagosome: defence against host stresses. *Cell Microbiol* **11**, 1170-8 (2009).

11. Akerley, B.J. et al. Systematic identification of essential genes by in vitro mariner mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America* **95** (10998).

12. Rubin, E.J. et al. In vivo transposition of mariner-based elements in enteric bacteria and mycobacteria. *Proceedings of the National Academy of Sciences of the United States of America* **96** (1999).

13. Sassetti, C.M. & Rubin, E.J. Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**, 12989-94 (2003).

14. Pettersson, E., Lundeberg, J. & Ahmadian, A. Generations of sequencing technologies. *Genomics* **93**, 105-11 (2009).

15. Griffin, J.E. et al. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* **7**, e1002251 (2011).

16. Goodman, A.L. et al. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**, 279-89 (2009).

17. Langridge, G.C. et al. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res* **19**, 2308-16 (2009).

18. Gawronski, J.D., Wong, S.M., Giannoukos, G., Ward, D.V. & Akerley, B.J. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc Natl Acad Sci U S A* **106**, 16422-7 (2009).

19. van Opinjen, T., Bodi, K.L. & Camilli, A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods* **6**, 767-772 (2009).

20. Flynn, J.L. et al. Tumor Necrosis Factor-alpha Is Required in the Protective Immune Response Against Mycobacteriumtuberculosis in Mice. *Immunity* **2**, 561-572 (1995).

21. Flynn, J.L. et al. An Essential Role for Interferon gamma in Resistance to Mycobacterium tuberculosis Infection. *Journal of Expermental Medicine* **178** (1993).

22. Flynn, J.L., Scanga, C.A., Tanaka, K.E. & Chan, J. Effects of Aminoguanidine on Latent Murine Tuberculosis. *Journal of immunology (Baltimore, Md : 1950)* **160**, 1796-1803 (1998).

23. Caruso, A.M. et al. Mice Deficient in CD4 T Cells Have Only Transiently Diminished Levels of IFN-{gamma}, Yet Succumb to Tuberculosis. *Journal of immunology (Baltimore, Md : 1950)* **162**, 5407-5416 (1999).

24. Mogues, T., Goodrich, M.E., Ryan, L., Lacourse, R. & North, R.J. The Relative Importance of T Cell Subsets in Immunity and Immunopathology of Airborne Mycobacterium tuberculosis Infection in Mice. *Journal of Experimental Medicine* **193**, 271-280 (2001).

25. Scanga, C.A. et al. Depletion of CD4+ T Cells Causes Reactivation of Murine Persistent Tuberculosis Despite Continued Expression of Interferon-gamma and Nitric Oxide Synthase 2. *Journal of Experimental Medicine* **192**, 347-358 (2000).

26. Macpherson, P. et al. Risk factors for mortality in smear-negative tuberculosis suspects: a cohort study in Harare, Zimbabwe. *Int J Tuberc Lung Dis* **15**, 1390-6 (2011).

27. Pawlowski, A., Jansson, M., Skold, M., Rottenberg, M.E. & Kallenius, G. Tuberculosis and HIV co-infection. *PLoS Pathog* **8**, e1002464 (2012).

28. Selwyn, P.A. et al. A Prospective Study of the Risk of Tuberculosis Among Intravenous Drug Users with Human Immunodeficiency Virus Infection. *The New England Journal of Medicine* **320**, 545-550 (1989).

29. Murray, H.W. et al. Role of tryptophan degradation in respiratory burst-independent antimicrobial activity of gamma interferon-stimulated human macrophages. *Infection and immunity* **57**, 845-849 (1989).

30. Byrne, G.I., Lehmann, L.K. & Landry, G.J. Induction of tryptophan catabolism is the mechanism for gamma-interferon-mediated inhibition of intracellular Chlamydia psittaci replication in T24 cells. *Infection and immunity* **53**, 347-351 (1986).

31. Beatty, W.L., Belanger, T.A., Desai, A.A., Morrison, R.P. & Byrne, G.I. Tryptophan depletion as a mechanism of gamma interferon-mediated chlamydial persistence. *Infection and immunity* **62**, 3705-3711 (1994).

32. Darwin, K.H., Ehrt, S., Gutierrez-Ramos, J.C., Weich, N. & Nathan, C.F. The proteasome of Mycobacterium tuberculosis is required for resistance to nitric oxide. *Science* **302**, 1963-6 (2003).

33. Darwin, K.H. & Nathan, C.F. Role for nucleotide excision repair in virulence of Mycobacterium tuberculosis. *Infect Immun* **73**, 4581-7 (2005).

34. Marrero, J., Rhee, K.Y., Schnappinger, D., Pethe, K. & Ehrt, S. Gluconeogenic carbon flow of tricarboxylic acid cycle intermediates is critical for Mycobacterium tuberculosis to establish and maintain infection. *Proc Natl Acad Sci U S A* **107**, 9819-24 (2010).

35. McKinney, J.D. et al. Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**, 735-738 (2000).

36. Munoz-Elias, E.J. & McKinney, J.D. Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nat Med* **11**, 638-44 (2005).

37. Pandey, A.K. & Sassetti, C.M. Mycobacterial persistence requires the utilization of host cholesterol. *Proc Natl Acad Sci U S A* **105**, 4376-80 (2008).

38. Murry, J.P., Pandey, A.K., Sassetti, C.M. & Rubin, E.J. Phthiocerol dimycocerosate transport is required for resisting interferon-gamma-independent immunity. *J Infect Dis* **200**, 774-82 (2009).

39. WHO. (http://www.who.int/tb/publications/global_report/2011/en/, 2011).

40. Valway, S.E. et al. Outbreak of Multi-Drug-resistant Tuberculosis in a New York State Prison, 1991. *American Journal of Epidemiology* **140**, 113-122 (1994).

41. Yerokhin, V.V., Punga, V.V. & Rybka, L.N. Tuberculosis in Russia and the Problem of Multiple Drug Resistance. *Annals of the New York Academy of Sciences* **953**, 133-137 (2006).

42. Li, T. et al. Impact of new migrant populations on the spatial distribution of tuberculosis in Beijing. *Int J Tuberc Lung Dis* **15**, 163-168 (2011).

43. Farmer, P. & Campos, N.G. Rethinking Medical Ethics: A View from Below. *Developing World Bioethics* **4**, 1471 (2004).